Cochrane Library

Cochrane Database of Systematic Reviews

Arthroscopic surgery for degenerative knee disease (osteoarthritis including degenerative meniscal tears) (Review)

O'Connor D, Johnston RV, Brignardello-Petersen R, Poolman RW, Cyril S, Vandvik PO, Buchbinder R

O'Connor D, Johnston RV, Brignardello-Petersen R, Poolman RW, Cyril S, Vandvik PO, Buchbinder R.
Arthroscopic surgery for degenerative knee disease (osteoarthritis including degenerative meniscal tears).
Cochrane Database of Systematic Reviews 2022, Issue 3. Art. No.: CD014328.
DOI: 10.1002/14651858.CD014328.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

ABSTRACT 1
PLAIN LANGUAGE SUMMARY 2
SUMMARY OF FINDINGS 5
BACKGROUND 8
OBJECTIVES 10
METHODS 10
RESULTS 14
Figure 1. 15
Figure 2. 19
Figure 3. 21
Figure 4. 23
DISCUSSION 29
AUTHORS' CONCLUSIONS 32
ACKNOWLEDGEMENTS 33
REFERENCES 34
CHARACTERISTICS OF STUDIES 41
DATA AND ANALYSES 110
Analysis 1.1. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 1: Pain (lower score=less pain) 112
Analysis 1.2. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 2: Function (higher score=better function) 113
Analysis 1.3. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 3: Knee-specific quality of life (higher score=better) 114
Analysis 1.4. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 4: Generic quality of life (higher score=better) 114
Analysis 1.5. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 5: Participant-reported success 115
Analysis 2.1. Comparison 2: Arthroscopic surgery versus exercise, Outcome 1: Pain (lower score=less pain) 117
Analysis 2.2. Comparison 2: Arthroscopic surgery versus exercise, Outcome 2: Function (higher score=better function) 118
Analysis 2.3. Comparison 2: Arthroscopic surgery versus exercise, Outcome 3: Knee-specific quality of life (higher score=better) 119
Analysis 2.4. Comparison 2: Arthroscopic surgery versus exercise, Outcome 4: Generic quality of life (higher score=better) 120
Analysis 2.5. Comparison 2: Arthroscopic surgery versus exercise, Outcome 5: Participant-reported success 120
Analysis 3.1. Comparison 3: Arthroscopic surgery versus glucocorticoid injection, Outcome 1: Function (OKS, 0-48, higher score=better function) 121
Analysis 4.1. Comparison 4: Arthroscopic surgery versus non-arthroscopic lavage, Outcome 1: Pain (AIMS-P subscale, 0-10, lower score=less pain) 122
Analysis 4.2. Comparison 4: Arthroscopic surgery versus non-arthroscopic lavage, Outcome 2: Function (AIMS-PF subscale, 0-10, higher score=better function) 122
Analysis 4.3. Comparison 4: Arthroscopic surgery versus non-arthroscopic lavage, Outcome 3: Participant-reported success (≥1cm improvement in VAS) 122
Analysis 5.1. Comparison 5: Arthroscopic surgery versus NSAIDs, Outcome 1: Participant-reported success 123
Analysis 6.1. Comparison 6: Arthroscopic surgery versus hyaluronic acid injections, Outcome 1: Pain (KSSS pain score of 30 or higher; higher=less pain) 124
Analysis 7.1. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 1: Serious adverse events 125
Analysis 7.2. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 2: Total adverse events 125
Analysis 7.3. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 3: Progression of knee osteoarthritis 126
Analysis 7.4. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 4: Subsequent knee surgery (replacement or high tibial osteotomy) 126
Analysis 8.1. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 1: Pain up to 3 months (lower score=less pain) 128
Analysis 8.2. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 2: Pain at >3 months up to 6 months (lower score=less pain) 128
Analysis 8.3. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 3: Pain at >6 months up to 2 years (lower score=less pain) 129

Analysis 8.4. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 4: Pain at >2 years up to 5 years (lower score=less pain) 129
Analysis 8.5. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 5: Function up to 3 months (higher score=better function) 130
Analysis 8.6. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 6: Function at >3 months up to 6 months (higher score=better function) 130
Analysis 8.7. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 7: Function at >6 months up to 2 years (higher score=better function) 131
Analysis 8.8. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 8: Function at >2 years up to 5 years (higher score=better function) 131
Analysis 9.1. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 1: Pain up to 3 months (lower score=less pain) 133
Analysis 9.2. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 2: Pain at >3 months up to 6 months (lower score=less pain) 133
Analysis 9.3. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 3: Pain at >6 months up to 2 years (lower score=less pain) 134
Analysis 9.4. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 4: Pain at >2 years up to 5 years (lower score=less pain) 134
Analysis 9.5. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 5: Function up to 3 months (higher score=better function) 135
Analysis 9.6. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 6: Function at >3 months up to 6 months (higher score=better function) 135
Analysis 9.7. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 7: Function at >6 months up to 2 years (higher score=better function) 136
Analysis 9.8. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 8: Function at >2 years up to 5 years (higher score=better function) 136
Analysis 10.1. Comparison 10: Sensitivity analysis: low risk of selection bias, Outcome 1: Pain up to 3 months (lower score=less pain) 137
Analysis 10.2. Comparison 10: Sensitivity analysis: low risk of selection bias, Outcome 2: Function up to 3 months (higher score=better function) 137
Analysis 11.1. Comparison 11: Sensitivity analysis: low risk of detection bias, Outcome 1: Pain up to 3 months (lower score=less pain) 138
Analysis 11.2. Comparison 11: Sensitivity analysis: low risk of detection bias, Outcome 2: Function up to 3 months (higher score=better function) 139
Analysis 12.1. Comparison 12: Sensitivity analysis: fixed-effect model, Outcome 1: Pain (lower score=less pain) 139
Analysis 12.2. Comparison 12: Sensitivity analysis: fixed-effect model, Outcome 2: Function (higher score=better function) 140
Analysis 13.1. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 1: Pain (lower score=less pain) 142
Analysis 13.2. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 2: Function (higher score=better function) 143
Analysis 13.3. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 3: Knee-specific quality of life (higher score=better) 144
Analysis 13.4. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 4: Generic quality of life (higher score=better) 145
Analysis 13.5. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 5: Participant-reported success 145
ADDITIONAL TABLES 145
APPENDICES 155
CONTRIBUTIONS OF AUTHORS 157
DECLARATIONS OF INTEREST 157
SOURCES OF SUPPORT 158
DIFFERENCES BETWEEN PROTOCOL AND REVIEW 158
INDEX TERMS 158

[Intervention Review]

Arthroscopic surgery for degenerative knee disease (osteoarthritis including degenerative meniscal tears)

Denise O'Connor[1], Renea V Johnston[1], Romina Brignardello-Petersen[2], Rudolf W Poolman[3], Sheila Cyril[1], Per O Vandvik[4], Rachelle Buchbinder[1]

[1]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University; Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology, Cabrini Health, Melbourne, Australia. [2]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. [3]Department of Orthopaedic Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam, Netherlands. [4]Department of Medicine, Lovisenberg Diaconal Hospital, Oslo, Norway

Contact: Rachelle Buchbinder, rachelle.buchbinder@monash.edu.

Editorial group: Cochrane Musculoskeletal Group.
Publication status and date: New, published in Issue 3, 2022.

Citation: O'Connor D, Johnston RV, Brignardello-Petersen R, Poolman RW, Cyril S, Vandvik PO, Buchbinder R.Arthroscopic surgery for degenerative knee disease (osteoarthritis including degenerative meniscal tears). *Cochrane Database of Systematic Reviews* 2022, Issue 3. Art. No.: CD014328. DOI: 10.1002/14651858.CD014328.

ABSTRACT

Background

Arthroscopic knee surgery remains a common treatment for symptomatic knee osteoarthritis, including for degenerative meniscal tears, despite guidelines strongly recommending against its use. This Cochrane Review is an update of a non-Cochrane systematic review published in 2017.

Objectives

To assess the benefits and harms of arthroscopic surgery, including debridement, partial menisectomy or both, compared with placebo surgery or non-surgical treatment in people with degenerative knee disease (osteoarthritis, degenerative meniscal tears, or both).

Search methods

We searched the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, Embase, and two trials registers up to 16 April 2021, unrestricted by language.

Selection criteria

We included randomised controlled trials (RCTs), or trials using quasi-randomised methods of participant allocation, comparing arthroscopic surgery with placebo surgery or non-surgical interventions (e.g. exercise, injections, non-arthroscopic lavage/irrigation, drug therapy, and supplements and complementary therapies) in people with symptomatic degenerative knee disease (osteoarthritis or degenerative meniscal tears or both). Major outcomes were pain, function, participant-reported treatment success, knee-specific quality of life, serious adverse events, total adverse events and knee surgery (replacement or osteotomy).

Data collection and analysis

Two review authors independently selected studies for inclusion, extracted data, and assessed risk of bias and the certainty of evidence using GRADE. The primary comparison was arthroscopic surgery compared to placebo surgery for outcomes that measured benefits of surgery, but we combined data from all control groups to assess harms and knee surgery (replacement or osteotomy).

Main results

Sixteen trials (2105 participants) met our inclusion criteria. The average age of participants ranged from 46 to 65 years, and 56% of participants were women. Four trials (380 participants) compared arthroscopic surgery to placebo surgery. For the remaining trials, arthroscopic surgery was compared to exercise (eight trials, 1371 participants), a single intra-articular glucocorticoid injection (one trial, 120 participants), non-arthroscopic lavage (one trial, 34 participants), non-steroidal anti-inflammatory drugs (one trial, 80 participants) and weekly hyaluronic acid injections for five weeks (one trial, 120 participants). The majority of trials without a placebo control were susceptible to bias: in particular, selection (56%), performance (75%), detection (75%), attrition (44%) and selective reporting (75%) biases. The placebo-controlled trials were less susceptible to bias and none were at risk of performance or detection bias. Here we limit reporting to the main comparison, arthroscopic surgery versus placebo surgery.

High-certainty evidence indicates arthroscopic surgery leads to little or no difference in pain or function at three months after surgery, moderate-certainty evidence indicates there is probably little or no improvement in knee-specific quality of life three months after surgery, and low-certainty evidence indicates arthroscopic surgery may lead to little or no difference in participant-reported success at up to five years, compared with placebo surgery.

Mean post-operative pain in the placebo group was 40.1 points on a 0 to 100 scale (where lower score indicates less pain) compared to 35.5 points in the arthroscopic surgery group, a difference of 4.6 points better (95% confidence interval (CI) 0.02 better to 9 better; I^2 = 0%; 4 trials, 309 participants). Mean post-operative function in the placebo group was 75.9 points on a 0 to 100 rating scale (where higher score indicates better function) compared to 76 points in the arthroscopic surgery group, a difference of 0.1 points better (95% CI 3.2 worse to 3.4 better; I^2 = 0%; 3 trials, 302 participants).

Mean post-operative knee-specific health-related quality of life in the placebo group was 69.7 points on a 0 to 100 rating scale (where higher score indicates better quality of life) compared with 75.3 points in the arthroscopic surgery group, a difference of 5.6 points better (95% CI 0.36 better to 10.68 better; I^2 = 0%; 2 trials, 188 participants). We downgraded this evidence to moderate certainty as the 95% confidence interval does not rule in or rule out a clinically important change.

After surgery, 74 out of 100 people reported treatment success with placebo and 82 out of 100 people reported treatment success with arthroscopic surgery at up to five years (risk ratio (RR) 1.11, 95% CI 0.66 to 1.86; I^2 = 53%; 3 trials, 189 participants). We downgraded this evidence to low certainty due to serious indirectness (diversity in definition and timing of outcome measurement) and serious imprecision (small number of events).

We are less certain if the risk of serious or total adverse events increased with arthroscopic surgery compared to placebo or non-surgical interventions. Serious adverse events were reported in 6 out of 100 people in the control groups and 8 out of 100 people in the arthroscopy groups from eight trials (RR 1.35, 95% CI 0.64 to 2.83; I^2 = 47%; 8 trials, 1206 participants). Fifteen out of 100 people reported adverse events with control interventions, and 17 out of 100 people with surgery at up to five years (RR 1.15, 95% CI 0.78 to 1.70; I^2 = 48%; 9 trials, 1326 participants). The certainty of the evidence was low, downgraded twice due to serious imprecision (small number of events) and possible reporting bias (incomplete reporting of outcome across studies). Serious adverse events included death, pulmonary embolism, acute myocardial infarction, deep vein thrombosis and deep infection.

Subsequent knee surgery (replacement or high tibial osteotomy) was reported in 2 out of 100 people in the control groups and 4 out of 100 people in the arthroscopy surgery groups at up to five years in four trials (RR 2.63, 95% CI 0.94 to 7.34; I^2 = 11%; 4 trials, 864 participants). The certainty of the evidence was low, downgraded twice due to the small number of events.

Authors' conclusions

Arthroscopic surgery provides little or no clinically important benefit in pain or function, probably does not provide clinically important benefits in knee-specific quality of life, and may not improve treatment success compared with a placebo procedure. It may lead to little or no difference, or a slight increase, in serious and total adverse events compared to control, but the evidence is of low certainty. Whether or not arthroscopic surgery results in slightly more subsequent knee surgery (replacement or osteotomy) compared to control remains unresolved.

PLAIN LANGUAGE SUMMARY

Arthroscopic surgery for degenerative knee disease

Background

Degenerative knee disease (osteoarthritis in the knee which affects the joint lining and menisci) is the most common cause of knee pain, swelling and stiffness in the knee joint which leads to difficulty in walking. The cartilage in the knee joint is damaged, resulting in friction in the joint surfaces and formation of new bone in severe cases. Arthroscopic knee surgery removes damaged cartilage and loose tissue and smooths the knee joint surfaces.

Study characteristics

We included 16 randomised trials (2105 participants) published up to 16 April 2021. Trials were conducted in Canada, Denmark, Finland, Italy, Norway, Pakistan, South Korea, Spain, Sweden, Netherlands and USA.

Overall, 56% of participants were women. The average age of participants ranged from 46 to 65 years and the average duration of symptoms ranged from 1.6 months to 4.4 years. Of the nine trials reporting their funding source, none received funding from industry. The other seven trials did not report any funding source.

We limit reporting to the main comparison, arthroscopic surgery versus placebo (dummy or sham) surgery.

Key results

Compared with placebo surgery, arthroscopic surgery had little benefit:

Pain (lower scores mean less pain)

Improvement in pain was 4.6 points better (0.02 better to 9 better) on a 0 to 100 point scale with arthroscopic surgery than with placebo, 3 months after surgery.

• People who had arthroscopic surgery rated their post-operative pain as 35.5 points.

• People who had placebo surgery rated their post-operative pain as 40.1 points.

Knee function (higher scores mean better function)

Improvement in knee function was 0.1 points better (3.2 worse to 3.4 better) on a 0 to 100 point scale with arthroscopic surgery than with placebo, 3 months after surgery.

• People who had arthroscopic surgery rated their post-operative knee function as 76.0 points.

• People who had placebo surgery rated their post-operative knee function as 75.9 points.

Knee-specific quality of life (higher scores mean better quality of life)

Improvement in knee-specific quality of life was 5.6 points better (0.4 better to 10.7 better) on a 0 to 100 point scale with arthroscopic surgery than with placebo, 3 months after surgery.

• People who had arthroscopic surgery rated their post-operative quality of life as 75.3 points.

• People who had placebo surgery rated their post-operative quality of life as 69.7 points.

Treatment success (rated by participants)

8% more people rated their treatment a success (25% fewer to 63% more), or 8 more people out of 100, at up to 5 years after surgery.

• 82 out of 100 people reported treatment success with arthroscopic surgery.

• 74 out of 100 people reported treatment success with placebo surgery.

Serious adverse events

2% more people (2% fewer to 10% more) had serious adverse events, or 2 more people out of 100, at up to 5 years after surgery.

• 8 out of 100 people reported serious adverse events with arthroscopic surgery.

• 6 out of 100 people reported serious adverse events with placebo surgery.

Total adverse events

2% more people (3% fewer to 11% more), had adverse events, or 2 more people out of 100, at up to 5 years after surgery.

• 17 out of 100 people reported adverse events with arthroscopic surgery.

• 15 out of 100 people reported adverse events with placebo surgery.

Subsequent knee surgery

2% more people (0.1% fewer to 9% more), had subsequent knee surgery, or 2 more people out of 100, at up to 5 years.

• 4 out of 100 people had knee replacement or osteotomy (knee surgery that reshapes bone) with arthroscopic surgery.

• 2 out of 100 people had knee replacement or osteotomy with placebo surgery.

Certainty of the evidence

We are confident that knee arthroscopy does not provide any clinically important benefits in terms of pain and function. We are moderately confident that knee arthroscopy probably does not provide any clinically important benefits in knee-specific quality of life over a placebo procedure. Knee arthroscopy may not increase participant-reported success compared with placebo. We have little confidence in the evidence because of differences across trials in reporting success and the small number of events. We are less certain of the risk of serious and total adverse events in arthroscopy versus placebo surgery: the evidence was uncertain because of the small number of events and incomplete reporting of study information.

Adverse events associated with surgery include total knee replacement, osteotomy, repeat arthroscopy, arthroscopy in opposite knee, cutaneous nerve lesion (damage to nerves in the skin), deep or superficial infection, general knee pain, swelling, instability, stiffness or decreased range of motion in the affected or opposite knee, haemarthrosis (bleeding into the knee joint), death, acute myocardial infarction (heart attack), hypoxaemia (decreased oxygen in the blood), deep vein thrombosis (blood clot in the deep veins), tendonitis (inflammation of tendons), pain from fall or other trauma, rupture of a Baker's cyst (a fluid-filled sac behind the knee), and back or hip or foot pain.

Arthroscopic surgery may or may not lead to slightly more subsequent knee surgery (replacement or osteotomy) than the placebo procedure.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings

Arthroscopic surgery compared to placebo for degenerative knee disease

Patient or population: people with degenerative knee disease (osteoarthritis including degenerative meniscal tears)
Setting: surgical
Intervention: arthroscopic surgery
Comparison: placebo for benefits and all control groups for adverse events including knee replacement

Outcomes	Anticipated absolute effects* (95% CI)			Relative effect (95% CI)	No of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Placebo	Arthroscopic surgery	Difference				
Pain [a] Scale: 0 to 100, 0 is no pain Follow-up: 3 months	The mean pain in the placebo group was **40.1 points**[b]	The mean pain in the arthroscopic surgery group was **35.5 points**	4.6 points better (0.02 better to 9 better)[c]		309 (4 studies)	⊕⊕⊕⊕ **High**[d]	SMD -0.23 (-0.45 to -0.001). Knee arthroscopic surgery results in little or no clinically important improvement in pain. Absolute change 5% better (0.02% better to 9% better) Relative change 8% better (0.03% better to 15% better)[e]
Knee function [a] Scale: 0 to 100, 100 is best function Follow-up: 3 months	The mean knee function in the placebo group was **75.9 points**[b]	The mean knee function in the arthroscopic surgery group was **76 points**	0.1 points better (3.2 worse to 3.4 better)		302 (3 studies)	⊕⊕⊕⊕ **High**[d]	SMD 0.01 (-0.22 to 0.23). Knee arthroscopic surgery results in little or no improvement in function. Absolute change 0.1% better (3% worse to 3% better) Relative change 0.2% better (5% worse to 6% better)[e]
Knee-specific health-related quality of life [a] Scale: 0 to 100, 100 is best quality of life Follow-up: 3 months	The mean quality of life in the placebo group was **69.7 points**[b]	The mean quality of life in the arthroscopic surgery group was **75.3 points**	5.6 points better (0.4 better to 10.7 better)		188 (2 studies)	⊕⊕⊕⊖ **Moderate**[f]	SMD 0.31 (0.02 to 0.59). Knee arthroscopic surgery probably provides little or no clinically important improvement in knee-specific quality of life.

							Absolute change 6% better (0.4% to 11% better). Relative change 11% better (0.8% better to 20% better)
Participant-reported success Last follow-up	74%	82% (49% to 100%)	8% more (25% fewer to 63% more)	RR 1.11 (0.66 to 1.86)	189 (3 studies)	⊕⊕⊖⊖ **Low**[f,g]	Knee arthroscopic surgery may result in little or no improvement in the number of people reporting success. Relative change 11% more reported success (34% fewer to 86% more)
Serious adverse events Last follow-up[i] Events include repeat arthroscopy, pulmonary embolism, deep vein thrombosis, heart attack, death, knee surgery, postoperative knee infection, anterior cruciate ligament reconstruction	5.6%	7.6% (3.6% to 15.8%)	2% more (2% fewer to 10% more)	RR 1.35 (0.64 to 2.83)	1206 (8 studies)[j]	⊕⊕⊖⊖ **Low**[f,h]	Knee arthroscopy may or may not lead to more serious adverse events. Relative change 35% more (36% fewer to 183% more)
Total adverse events Last follow-up[i] Events include serious events and less serious transient pain in the back, hip, foot, tendonitis, syncope, rupture of Baker's cyst, pain and swelling in index knee after surgery, superficial infection, haemarthrosis, cutaneous nerve lesion, nausea, dizziness	15.0%	17.2% (11.7% to 25.5%)	2% more (3% fewer to 11% more)	RR 1.15 (0.78 to 1.70)	1326 (9 studies)[j]	⊕⊕⊖⊖ **Low**[f,h]	Knee arthroscopy may or may not slightly increase total adverse events. Relative change 15% more (22% fewer to 70% more)
Knee surgery (replacement or osteotomy) Last follow-up[i]	1.5%	5% (1.4% to 10.8%)	2% more (0.1% fewer to 9% more)	RR 2.63 (0.94 to 7.34)	864 (4 studies)[j]	⊕⊕⊕⊖ **Low**[k]	Knee arthroscopy may or may not lead to slightly more knee surgery. Relative change 163% more (6% fewer to 634% more)

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: confidence interval; **RR:** risk ratio

GRADE Working Group grades of evidence

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

[a]Pain measured onnumerical rating scale (Sihvonen 2013), Knee-Specific Pain Scale (KSPS) (Moseley 2002), questionnaire designed specifically for the trial (Moseley 1996); Knee Injury and Osteoarthritis Outcome Score (KOOS) pain subscale (Roos 2018). Knee function measured on Knee Injury and Osteoarthritis Outcome Score (KOOS) (Roos 2018), Lysholm knee score (Sihvonen 2013), Short Form 36-item questionnaire (SF-36) bodily pain (Moseley 2002). Knee-related quality of life (QoL) measured on the KOOS Knee-related QoL subscale (Roos 2018), and the Western Ontario Meniscal Evaluation tool (WOMET) (Sihvonen 2013)

[b]Control group risk was estimated from the placebo value at follow-up for pain, knee function and knee-related quality of life in Sihvonen 2013

[c]Standardised mean difference (SMD) back-translated to typical scales by multiplying the SMD by the standard deviation (SD) at baseline in the placebo group as reported in Sihvonen 2013: mean (SD) for knee pain (0 to 100 scale): 60.1 (20.0); mean (SD) knee function (0 to 100 scale): 60.1 (14.6); mean (SD) generic quality of life (15D): 0.90 (0.06); mean (SD) knee-specific quality of life (WOMET 0 to 100 scale): 52.8 (18.1).

[d]Overall, the certainty of evidence was high at 3-month follow-up for pain and function. One trial measuring pain was at potential risk of selection bias, but this probably did not change our confidence in the effect estimates. The 95% confidence intervals exclude a clinically important change (defined as 12 points (minimum, maximum: 2, 30) on a 0 to 100 point pain scale; and 13 (3, 34) on a 0 to 100 point WOMAC function scale). Further research is likely to strengthen the conclusion that there was no important differences in pain and function between groups, rather than change the conclusion

[e]Relative change: absolute change (mean difference) divided by mean at baseline in the placebo group (values were: 60.1 points on 0 to 100 point pain scale; 60.1 points on 0 to 100 knee function scale; and 52.8 points on 0 to 100 quality of life scale; from Sihvonen 2013).

[f]Downgraded due to imprecision: the 95% confidence intervals do not rule in or rule out a clinically important change (defined as 10 points on the 0 to 100 point quality of life scale); or for dichotomous outcomes the total number of participants was small, or number of events was small (< 200); or data were from a single trial only.

[g]Downgraded due to indirectness for participant-reported success as there was diversity in definition and timing of measurement: reported at 6 months, 24 months and 5 years across trials.

[h]Downgraded due to possible reporting bias: incomplete reporting of outcome across studies.

[i]Total and serious adverse events were reported at 24 months (Roos 2018; Van de Graaf 2018); 25 months (Merchan 1993); and 5 years (Gauffin 2014; Herrlin 2007; Katz 2013; Kise 2016; Sihvonen 2013). Total adverse events only were reported at 6 months in one study (Saeed 2015).

[j]For serious adverse events, adverse events and subsequent knee surgery (replacement or osteotomy), we included trials that compared arthroscopy to placebo or to non-surgical interventions. For serious adverse events, comparison groups were placebo in 2 trials, exercise in 5 trials and oral non-steroidal anti-inflammatory drugs (NSAIDs) in 1 trial. For adverse events, the comparison groups included placebo in 2 trials, exercise in 5 trials, and oral NSAIDs and hyaluronic acid injections in single trials. For knee surgery, the comparison groups included placebo in 1 trial and exercise in 3 trials.

[k]Downgraded twice due to imprecision: the 95% confidence intervals do not rule in or rule out a clinically important change as the total number of events was small (< 200).

BACKGROUND

Description of the condition

Degenerative knee disease (osteoarthritis of the knee, which can include the joint lining, menisci, or both) is a prevalent musculoskeletal condition and a major contributor to disability globally. Its prevalence continues to increase with population growth, ageing and rising global obesity. It affects one or more compartments of the knee joint and its periarticular structures, including the articular cartilage, menisci, underlying bone and synovial lining. In the USA, it is estimated that about 25% of people aged 45 years or older have symptoms of degenerative knee disease that impact on their quality of life (Alkan 2014; Mahir 2016), while about 10% of the world's population aged 60 years or older have symptomatic osteoarthritis (Zhang 2010).

The major symptoms of knee osteoarthritis are pain, stiffness and swelling, which in turn can lead to impaired mobility and physical function. Symptoms commonly fluctuate, and may or may not progressively worsen. A symptomatic meniscal tear may be suggested by new onset of knee pain accompanied by mechanical symptoms, such as locking or catching, with medial or lateral joint line pain. The presence of a collection of symptoms, including localised pain, clicking, catching and giving way or buckling, has been found to more likely indicate the presence of a symptomatic meniscal tear on the basis of orthopaedic surgeon and magnetic resonance imaging (MRI) assessment, compared to the absence of all of these symptoms (Niu 2011). However, these symptoms are non-specific, and there is no current consensus on what defines a symptomatic meniscal tear (Buchbinder 2015). Similarly, signs such as medial or lateral joint line tenderness and loss of full extension, and named physical tests such as the McMurray test, purported to be useful in making a diagnosis of a symptomatic meniscal tear, have been found to have limited diagnostic accuracy (Hegedus 2007).

International guidelines and clinical care standards recommend that the diagnosis of degenerative knee disease be made on the basis of clinical features alone, unless an alternative diagnosis is suspected (ACQSHC 2017; NICE 2014; RACGP 2018; Zhang 2009). For example, the UK's National Institute for Health and Care Excellence (NICE) 2014 clinical guideline recommends that osteoarthritis be diagnosed clinically without investigations if a person is over 45 years and has activity-related joint pain and either has no morning joint-related stiffness or morning stiffness that lasts no longer than 30 minutes (NICE 2014).

When there is a suspicion of an alternate diagnosis (e.g. malignancy, insufficiency fracture), plain radiographs are the first line investigation in routine care (ACQSHC 2017), followed by MRI if there is continued suspicion of serious pathology not detected by X-ray. Weight-bearing X-rays are also indicated for severe symptoms that have not responded to non-operative treatment when contemplating joint replacement.

The Kellgren and Lawrence classification is widely used to grade the severity of osteoarthritis (Kellgren 2000). It includes five grades, ranging from 0 to 4, based on the increasing severity of osteoarthritis, where grade 0 indicates no osteoarthritis and grade 4 indicates severe osteoarthritis (Kellgren 2000).

MRI findings of degenerative knee disease are common in asymptomatic people and may be present even when plain radiographs are normal (Englund 2008). For example, one population-based study that included 710 people under 50 years old who had no radiographic evidence of knee osteoarthritis found the presence of at least one type of abnormality in 90% to 97% of those with symptoms and 86% to 88% of those with painless knees (Guermazi 2012). In a related study, one or more meniscal tears was present in 23% of those without symptoms, 32% in people with symptoms and 24% of those with no or equivocal radiographic evidence of osteoarthritis (Englund 2008). Most often it is the medial meniscus that is torn, and multiple tears are present in more than a third of patients (Bhattacharyya 2003; Englund 2008).

Description of the intervention

Currently, treatment of degenerative knee disease can be grouped broadly into two types: non-operative and operative management. Non-operative non-pharmacologic treatment includes achieving and maintaining a normal weight (Christensen 2007), and supervised land-based or aquatic exercise (Bartel 2016; Fransen 2015). Simple analgesia provides very minimal benefit (Leopoldino 2019), while non-steroidal anti-inflammatory drugs provide limited benefits (Puljak 2017), and may only be suitable for a subset of people without comorbidities that preclude their use. Intra-articular glucocorticoid injections may provide short-term pain relief (Jüni 2015). Joint replacement surgery, which is the only definitive treatment, is reserved for people with severe disease who have failed non-operative management (Brignardello-Petersen 2017).

Arthroscopic surgery is the most widely performed surgical procedure for degenerative knee disease. It involves the insertion of an arthroscope into the knee joint under either local or general anaesthesia through two portals on the front of the knee joint (Kise 2016). The two most frequently used arthroscopic procedures for managing degenerative knee disease are debridement and partial meniscectomy (Katz 2013). Arthroscopic debridement involves the removal of damaged cartilage, irrigation of the knee joint to wash out all debris, including cartilage fragments and loose bodies, and smoothing of the joint surfaces (Kirkley 2008). Arthroscopic partial meniscectomy involves the removal of torn meniscal fragments and trimming the meniscus back to a stable rim (Sihvonen 2013).

How the intervention might work

The overall mechanism of action of arthroscopic surgery is hypothesised to be via identification and removal of the mechanical components that contribute to the symptoms of osteoarthritis, such as damaged cartilage and loose bodies, while preserving the knee joint (Howell 2014; Kirkley 2008; Shin 2012). Theoretically, this process would reduce inflammation of the joint lining and improve joint motion, resulting in decreased pain and improved knee function (Mounsey 2009). Partial debridement of the torn meniscus is also hypothesised to lead to improvements in pain and mechanical symptoms (Howell 2014; Steadman 2007).

Why it is important to do this review

Evidence from randomised controlled trials that have included a placebo or exercise control, accumulating over two decades, indicates that arthroscopic surgery may provide limited benefit for people with degenerative knee disease, irrespective of osteoarthritis grade or the presence or absence of meniscal tears

(Moseley 2002; Herrlin 2007; Kirkley 2008; Katz 2013; Sihvonen 2013; Kise 2016).

Since 2013, evidence-based guidelines have consistently recommended against the use of arthroscopic debridement and lavage for symptomatic osteoarthritis of the knee, but have been inconsistent in their recommendations regarding treatment of degenerative meniscal tears (Australian Knee Society 2016; Brown 2013; McAlindon 2014; NICE 2014). Both the 2013 and updated 2015 second edition of the American Academy of Orthopaedic Surgeons (AAOS) guidelines for the treatment of knee osteoarthritis made a strong recommendation against performing arthroscopy with lavage, debridement, or both, in people with a primary diagnosis of symptomatic osteoarthritis of the knee (Brown 2013). However, the 2015 guidelines made an inconclusive recommendation regarding arthroscopic partial meniscectomy, stating that the lack of compelling evidence has resulted in an unclear balance between benefits and potential harm. The 2014 NICE clinical guideline for the care and management of osteoarthritis recommends against referral for arthroscopic lavage and debridement as part of treatment for osteoarthritis, "unless the person has knee osteoarthritis with a clear history of mechanical locking (as opposed to morning joint stiffness), 'giving way' or X-ray evidence of loose bodies"; meniscal tears are not specifically mentioned (NICE 2014). The 2014 Osteoarthritis Research Society International (OARSI) guideline does not make any recommendation regarding arthroscopy, citing the consistent evidence of ineffectiveness (McAlindon 2014). The 2016 Australian Knee Society Position Statement also indicates that arthroscopic debridement or lavage, or both, are not indicated as a primary treatment in the management of knee osteoarthritis, but indicates arthroscopy may be appropriate for symptomatic meniscal tears that have failed an appropriate trial of a structured rehabilitation program (Australian Knee Society 2016).

In contrast to previous guidelines, the updated 2018 Royal Australian College of General Practitioners (RACGP) guideline for the management of knee and hip osteoarthritis recommends against offering meniscectomy and cartilage repair for people with knee osteoarthritis unless the person also has mechanical symptoms of a clinically locked knee (RACGP 2018). This is also consistent with the 2017 Australian Osteoarthritis of the Knee Clinical Care Standard, which indicates that arthroscopic procedures should only be offered if the individual has true mechanical locking or another appropriate indication for these procedures (e.g. septic arthritis or as an investigation when MRI is not possible), on the basis that these treatments are ineffective (ACQSHC 2017).

In 2017, we published a Clinical Practice Guideline as part of the BMJ Rapid Recommendation series, in which we made a strong recommendation against the use of arthroscopy in nearly all people with degenerative knee disease, including those with or without imaging evidence of osteoarthritis, mechanical symptoms, or sudden symptom onset (Siemieniuk 2017). Triggered by a 2016 randomised controlled trial that found that knee arthroscopy was no better than exercise therapy for treating people with a degenerative medial meniscus tear (Kise 2016), the recommendation was based upon a systematic review that included 13 randomised controlled trials (1688 participants) with placebo and non-operative care controls that assessed benefits and harms, and 12 observational studies (> 1.8 million participants) that also contributed to assessment of potential harms (Brignardello-Petersen 2017). The review identified:

- high-certainty evidence that knee arthroscopy provides a very small reduction in pain up to three months (mean difference (MD) = 5.4 on a 100-point scale, 95% CI 2.0 to 8.8), and very small or no pain reduction up to two years (MD = 3.1, 95% CI -0.2 to 6.4); and
- moderate-certainty evidence that knee arthroscopy results in a very small improvement in function in the short term (MD = 4.9 on a 100-point scale, 95% CI 1.5 to 8.4) and very small or no improvement in function up to two years (MD = 3.2, 95% CI -0.5 to 6.8)

when compared to conservative management (which included various controls, including placebo surgery). There was low-certainty evidence of a very low probability of serious complications after knee arthroscopy.

Despite the availability of evidence of a lack of clinically relevant benefit from arthroscopic partial meniscectomy over placebo surgery or a structured exercise program, there continues to be a lack of consensus among orthopaedic surgeons regarding the current place of this procedure in the routine management of people with knee symptoms putatively attributed to a degenerative meniscal tear (Lohmander 2019). Recently, the British Association for Surgery of the Knee (BASK) Meniscal Consensus Project published a guideline specifically focused on meniscal tears (Abram 2019a). Informed by evidence and based upon a consensus approach to management of common clinical presentations, it recommended against arthroscopic meniscal surgery in people with advanced osteoarthritis, except in rare special cases, but recommended offering it to people with 'meniscal' or 'possible meniscal' symptoms and signs and a 'meniscal target' who fail to respond to a period of non-operative treatment. Earlier surgery could also be considered if deemed appropriate by an experienced colleague acting as a second opinion. However, a recent post hoc analysis of a randomised controlled trial that found no benefit of partial meniscectomy over placebo failed to identify a subgroup who might benefit (Sihvonen 2018), while others have failed to identify specific patient characteristics that might predict a more favourable outcome following meniscal surgery (Pihl 2020).

While our review only considered potential harms of arthroscopy up to three months post surgery, accumulating evidence from both observational studies and randomised controlled trials suggest that arthroscopic partial meniscectomy may be associated with worsening of the underlying osteoarthritis, accompanied by an increased risk of joint replacement surgery, particularly in older people (Abram 2019b; Dearing 2010; Harris 2013; Hawker 2008; Katz 2013; Roemer 2017; Wai 2002).

Synthesis of all the available evidence is therefore warranted to determine the balance of benefits to harms of arthroscopic surgery for degenerative knee disease. This Cochrane Review is an update of an earlier systematic review (Brignardello-Petersen 2017). The updated review has been conducted according to the guidelines recommended by the Cochrane Musculoskeletal Editorial Board (Ghogomu 2014). The Cochrane format emphasises assessment of placebo-controlled trials separately from other controls, to enable more discriminating estimates of benefits of arthroscopic surgery per se, from any differences compared to other treatments, and to determine if there are any differences in outcomes between

those with meniscal tear and those without. Other updates to the methods are described where relevant.

OBJECTIVES

To assess the benefits and harms of arthroscopic surgery, including debridement, partial menisectomy or both, compared with placebo surgery or non-surgical treatment in people with degenerative knee disease (osteoarthritis, degenerative meniscal tears, or both).

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials (RCTs), or trials using quasi-randomised methods of participant allocation. Studies reported as full-text, those published as abstract only, and unpublished data were all eligible for inclusion. There were no language or date restrictions. In contrast to our original review (Brignardello-Petersen 2017), which excluded trials with fewer than 10 participants, we did not exclude trials based upon their size. We also limited our review of harms to trial data.

Types of participants

We included people with symptomatic (defined as persistent knee pain that affects quality of life) degenerative knee disease (osteoarthritis, degenerative meniscal tears, or both). There was no age limit.

We excluded participants with acute traumatic knee pain.

Types of interventions

We included trials comparing arthroscopic surgery that included debridement or partial meniscectomy, or both, with:

- placebo surgery (primary comparison as it is least prone to bias); and
- non-surgical interventions, including: exercise and other physical therapy interventions; injections (including glucocorticoid injection, platelet-rich plasma or cell-based therapies such as stem cell therapy); non-arthroscopic lavage/irrigation, drug therapy (including simple analgesia and non-steroidal anti-inflammatory drugs); and supplements and complementary therapies.

We excluded arthroscopic joint lavage alone as this intervention is covered by a separate Cochrane Review (Reichenbach 2010), as is osteotomy (Brouwer 2014). We also excluded studies that compared one type of arthroscopic procedure to another type, or to another type of surgery.

Co-interventions were allowed, provided they were applied equally in all treatment groups.

Types of outcome measures

Major outcomes

- Overall pain measured on a visual analogue scale (VAS), numerical rating scale or other scales, or pain subscales of composite scales if separate pain scales were not reported.
- Function measured on a visual analogue scale (VAS), numerical rating scale or other scales, or function or activities of daily living scales (e.g. Lysholm Knee Scoring Scale) or subscales of composite scales (e.g. Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) physical function scale, or Knee Injury and Osteoarthritis Outcome Score (KOOS) Activities of Daily Living (ADL) subscale), if separate function scales were not reported.
- Knee-specific health-related quality of life using scales such as the Western Ontario Meniscal Evaluation Tool (WOMET) if available.
- Participant-reported treatment success as defined by the trialists.
- Serious adverse events: proportion in each group with serious adverse events, defined as an event that leads to hospitalisation, disability or death (such as deep vein thrombosis, cardiovascular or pulmonary events and including knee surgery).
- Total adverse events: proportion in each group experiencing any adverse event, mild or serious in nature, including deep vein thrombosis, infections, cardiovascular events and pulmonary embolism.
- Knee surgery (replacement or osteotomy): proportion in each group who subsequently had a knee replacement or osteotomy.

Minor outcomes

- Generic health-related quality of life (HRQoL) measured on a generic scale (e.g. SF-36 (36-item Short Form Health Survey); EQ-5D (EuroQoL 5-dimension instruments); 15D (a 15-dimensional, self-administered HRQoL instrument).
- Progression of knee osteoarthritis as defined by the trialists (e.g. Kellgren-Lawrence classification (Kellgren 1957), or Ahlback classification (Ahlback 1968)).

Time points

We stratified the analysis for pain, function and health-related quality of life by these follow-up time frames:

- up to three months;
- between three and six months;
- between six months and two years;
- between two and five years;
- between five and 10 years.

For participant-reported treatment success, serious adverse events, total adverse events, progression of knee osteoarthritis, and subsequent knee surgery (replacement or osteotomy), we extracted and reported events at the final follow-up.

Search methods for identification of studies

Electronic searches

We searched the following electronic databases on 16 April 2021.

- The Cochrane Central Register of Controlled Trials (CENTRAL) (via Ovid EBM Reviews, 16 April 2021).
- MEDLINE (via Ovid from 1946 to 16 April 2021).
- Embase (via Ovid from 1947 to 16 April 2021).

The search strategies are shown in Appendix 1, Appendix 2 and Appendix 3.

We also searched ClinicalTrials.gov (www.clinicaltrials.gov/) and the World Health Organization (WHO) trials portal (www.who.int/clinical-trials-registry-platform) for ongoing studies, using the terms 'arthroscopic' or 'arthroscopy' or 'debridement' and 'knee osteoarthritis' or 'meniscal degeneration' on 16 April 2021 (see Appendix 4).

No language restrictions were applied.

Searching other resources

We checked reference lists of all primary studies and review articles for additional references. We also searched Scopus (www.elsevier.com/solutions/scopus) for subsequent publications relating to the included trials on 16 April 2021.

Data collection and analysis

Selection of studies

Two review authors (RJ, SC) independently screened the titles and abstracts of the studies identified from the search and coded them as 'retrieve' (eligible or potentially eligible/unclear) or 'do not retrieve'. Two review authors (RJ, SC) independently screened the full-text versions of potentially eligible records and identified studies for inclusion, and identified and recorded reasons for exclusion of the ineligible studies. We resolved any disagreement through discussion or consulted a third author (DOC). We identified and excluded duplicates and collated multiple reports of the same study so that each study, rather than each report, was the unit of interest in the review. We completed a PRISMA flow diagram and 'Characteristics of excluded studies' table.

Data extraction and management

Two review authors (DOC, SC) extracted study characteristics from included studies. We extracted the following study characteristics.

- Methods: study design, setting, total duration of study, details of any 'run in' period, number of study centres and location, withdrawals and date of study.
- Participants: N, mean (SD) age, age range, sex, disease duration, severity of condition, diagnostic criteria, important condition-specific baseline data, inclusion criteria and exclusion criteria.
- Interventions: intervention, comparison, concomitant medications and excluded medications.
- Outcomes: primary and secondary outcomes specified and collected, and time points reported.
- Characteristics of the design of the trial, as outlined below in the Assessment of risk of bias in included studies section.
- Notes: trial registration, funding for trial, and notable declarations of interest of trial authors.

Two review authors (DOC, SC) extracted outcome data from included studies, including the number of events and number of participants per treatment group for dichotomous outcomes, and means and standard deviations and number of participants per treatment group for continuous outcomes. We noted in the 'Characteristics of included studies' tables if outcome data were not reported in a usable way and when data were transformed or estimated from a graph. We resolved disagreements by consensus or by involving a third author (RB). One review author (DOC) transferred data into Review Manager (Review Manager 2020). A second author (RJ) double-checked that data were entered correctly by comparing the data presented in the systematic review with the study reports.

In keeping with our previous review (Brignardello-Petersen 2017), where multiple measures of pain were reported, we extracted the measure highest on the following hierarchy recommended by Juhl 2012: (1) Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale; (2) pain during activity (visual analogue scale (VAS)); (3) pain during walking (VAS); (4) general knee pain (VAS); (5) pain at rest (VAS); (6) other composite pain scales (e.g. SF-36 bodily pain subscale, Arthritis Impact Measurement Scale (AIMS) pain subscale); and (7) other single item measures. The Knee Injury and Osteoarthritis Outcome Score (KOOS) pain subscale is considered equivalent to the WOMAC pain subscale. Where pain subscales of composite scales were not presented, we did not use total scores.

Similarly, where multiple measures of overall function were reported, we extracted the measure highest on the hierarchy recommended by Juhl 2012: (1) WOMAC function subscale; (2) SF-36 physical function subscale; (3) SF-36 (physical composite score); and (4) other composite disability scores. The KOOS ADL subscale is considered equivalent to the WOMAC function subscale. Where function subscales of composite scales were not presented, we used the total scores for function.

To prevent selective inclusion of data based on the results, we used the following a priori defined decision rules to select data from trials.

- Where trials reported outcomes at multiple time points, we extracted data from the latest time point within the period of time we were interested in.
- Where trialists reported both final values and change from baseline values for the same outcome, we extracted final values.
- Where trialists reported both unadjusted and adjusted values for the same outcome, we extracted unadjusted values.
- We extracted intention-to-treat (ITT)-analysed data for outcomes assessing benefits (pain, function, knee-specific and generic quality of life and participant-reported treatment success) if reported, or extracted the number of participants analysed at that time point, if data were not available for missing participants. For outcomes assessing harms (serious adverse events, total adverse events, progression of knee osteoarthritis, knee surgery (replacement or osteotomy)), we extracted data for those randomised to, and receiving, allocated treatment.

The primary comparison was arthroscopic surgery versus placebo surgery for outcomes that measured benefits of surgery (pain, function and health-related quality of life at three months, and treatment success at last follow-up), but we combined data from all control groups to assess harms (serious adverse events, total adverse events, progression of knee osteoarthritis, knee surgery (replacement or osteotomy)) at last follow-up.

Assessment of risk of bias in included studies

Two review authors (DOC, SC) independently assessed risk of bias for each included study using the criteria outlined in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2017).

Any disagreements were resolved through discussion or by involving another review author (RJ or RB). We assessed the risk of bias according to the following domains.

- Random sequence generation.
- Allocation concealment.
- Blinding of participants and personnel.
- Blinding of outcome assessment.
- Incomplete outcome data.
- Selective outcome reporting.
- Other bias: unexplained baseline imbalance (i.e. not explained by suboptimal randomisation), unit of analysis issues, inappropriate or unequal application of co-interventions across treatment groups, whether the number of cross-overs from the control group to arthroscopic surgery group biased the analysis.

We graded each potential source of bias as high, low or unclear risk, and gave a justification for our judgment, in the risk of bias tables. We summarised the risk of bias judgements across different studies for each of the domains listed. We considered blinding separately for different key outcomes (e.g. self-reported outcomes such as pain, function, participant-reported treatment success, health-related quality of life; and objective outcomes such as adverse events and knee surgery (replacement or osteotomy)). We considered the impact of missing data by key outcomes, where possible.

Where information on risk of bias relates to unpublished data or correspondence with a trialist, we noted this in the risk of bias table.

When considering treatment effects, we took into account the risk of bias for the studies that contributed to that outcome.

We present the figures generated by the risk of bias tool to provide summary assessments of the risk of bias.

Assessment of bias in conducting the systematic review

We conducted the review according to our prior published review (Brignardello-Petersen 2017), reporting any deviations from it in the Differences between protocol and review section of the systematic review.

Measures of treatment effect

We used the Cochrane Collaboration statistical software, Review Manager 5.4 (Review Manager 2020), to perform data analysis. We analysed dichotomous data as risk ratios (RR), with 95% confidence intervals (CI). Continuous data were analysed as mean difference (MD) or standardised mean difference (SMD), depending on whether the same scale was used to measure an outcome, with 95% CI. We entered data presented as a scale with a consistent direction of effect across studies.

When different scales were used to measure the same conceptual outcome (e.g. function), we calculated SMD, with corresponding 95% CIs. SMDs were back-translated to a typical scale by multiplying the SMD by a typical among-person standard deviation (SD) (e.g. the standard deviation of the control group at baseline from the most representative trial (Higgins 2021)). This contrasts with our original review which converted pain and function to 0 to 100 scales before meta-analysis, then calculated MD. For pain, we converted pooled SMD results to a 0 to 100 scale using the SD of 20 from Sihvonen 2013. For function, we converted SMD results to the KOOS 0 to 100 scale using the SD of 14.6 from Sihvonen 2013. For knee-specific health-related quality of life, we converted SMD results to the WOMET 0 to 100 scale using the SD of 18.1 from Sihvonen 2013. For generic health-related quality of life, we converted SMD results to the 15D 0 to 1 scale using the SD of 0.06 from Sihvonen 2013 (and also the SF-36 Mental Component Summary (MCS) score 0 to 100 scale using SD of 10 from Roos 2018).

In the 'Comments' column of the summary of findings table, we reported the absolute percent difference and the relative percent change from baseline. The number needed to treat for an additional beneficial outcome (NNTB) or number needed to treat for an additional harmful outcome (NNTH) will only be reported when the outcome shows a clinically important difference between treatment groups.

For dichotomous outcomes (treatment success, adverse events, progression of knee osteoarthritis, knee replacement or osteotomy), we calculated NNTB or NNTH from the control group event rate and the relative risk using the Visual Rx NNT calculator (Cates 2008). We calculated the NNTB for continuous measures (pain, function, health-related quality of life) using the Wells calculator (available at the Cochrane Musculoskeletal Group (CMSG) Editorial office, musculoskeletal.cochrane.org/).

In keeping with our previous review (Brignardello-Petersen 2017), we used the minimal important differences (MIDs) for pain, function and health-related quality of life from a linked systematic review performed to establish the most credible MIDs for each of the instruments used to measure these outcomes (Devji 2017). The most credible MID was the median of all the credible MIDs. For pain, we used the MID for WOMAC pain which was found to be 12 points (minimum, maximum: 2, 30) on a 0 to 100 point scale (noting that the MID for KOOS pain was also 12 (4, 20) also on a 0 to 100 point scale). For function, we used WOMAC function MID of 13 (3, 34) on a 0 to 100 point scale (noting that the MID for KOOS ADL was 8 (3, 9) also on a 0 to 100 point scale). For health-related quality of life, we used the MID for the EQ-5D which was 0.15 (minimum and maximum were not reported) on a -0.59 to 1 scale.

For dichotomous outcomes, we calculated the absolute percent change from the difference in the risks between the intervention and control group using GRADEpro (GRADEPro GDT) and expressed this as a percentage. The relative percent change was calculated as the risk ratio − 1 and expressed as a percentage. For continuous outcomes, we calculated the absolute benefit as the improvement in the intervention group minus the improvement in the control group (mean difference), in the original units, and expressed this as a percentage.

The relative percent change for dichotomous data was calculated as the Risk Ratio - 1 and expressed as a percentage. For continuous outcomes, the relative difference in the change from baseline was calculated as the absolute benefit divided by the baseline mean of the control group, expressed as a percentage.

Unit of analysis issues

The participant was the unit of analysis wherever possible. If a trial randomised participants to different treatment groups, but treated two knees in a single participant without adjusting for the lack of independence in the analysis, we reported this as a potential source

of other bias. We assessed the impact of including these trials in a sensitivity analysis. When the data for these studies were extracted, the number of knees was taken as the population for the study.

Dealing with missing data

When required, we contacted trial authors to obtain data missing from the trial reports. For outcomes assessing benefit (pain, function, knee-specific and generic health-related quality of life and participant-reported treatment success), we used the number of participants per group analysed at that time point. If the number of participants per group analysed was not presented for each time point, the number of randomised participants in each group at baseline was used. For outcomes assessing harms (severe adverse events, total adverse events, progression of knee osteoarthritis, knee surgery (replacement or osteotomy)), we used the number of participants receiving the allocated intervention as the denominator.

Where possible, we calculated missing standard deviations from other statistics, such as standard errors, confidence intervals or P values, according to the methods recommended in Chapter 6 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2021). If we could not calculate standard deviations, we imputed them from other studies in the meta-analysis, as per Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2021). Where data were calculated or imputed, we reported this in the notes section of the Characteristics of included studies.

Assessment of heterogeneity

We assessed the clinical and methodological diversity of the included studies, in terms of participants, interventions, outcomes and study characteristics, to determine whether a meta-analysis was appropriate. We assessed statistical heterogeneity by visually inspecting forest plots to check for obvious differences in results between the studies, and by using the I^2 and Chi^2 tests.

As recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2021), we interpreted an I^2 statistic for heterogeneity of 0% to 40% as 'might not be important'; 30% to 60% may represent 'moderate' heterogeneity; 50% to 90% may represent 'substantial' heterogeneity; and 75% to 100% represents 'considerable' heterogeneity. As noted in the *Cochrane Handbook for Systematic Reviews of Interventions*, the importance of I^2 depends on (i) the magnitude and direction of effects, and (ii) the strength of evidence for heterogeneity.

We interpreted the Chi^2 test so that $P \leq 0.10$ indicates evidence of statistical heterogeneity. Where we identified substantial heterogeneity, we reported it and investigated possible causes by following the recommendations in section 10 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2021).

Assessment of reporting biases

To assess small study effects, we planned to generate funnel plots for meta-analyses including at least 10 trials of varying size. If we detected asymmetry in the funnel plot, we planned to review the characteristics of the trials to assess whether the asymmetry was likely due to publication bias or other factors, such as the methodological or clinical diversity of the trials. Where we were able to pool more than 10 trials, we conducted formal statistical tests to investigate funnel plot asymmetry, and followed the recommendations in section 13 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Page 2021).

To assess outcome reporting bias, we checked trial protocols against published reports. For studies published after 1 July 2005, we searched the Clinical Trial Register at the International Clinical Trials Registry Platform of the World Health Organization (www.who.int/clinical-trials-registry-platform) for the a priori trial protocol. If trial protocols were unavailable, we compared the outcomes reported in the methods and results sections of the trial reports.

Data synthesis

We pooled outcomes that assessed the benefits of treatment across studies with a similar comparator and stratified the primary analysis by follow-up time, as follows:

- arthroscopic surgery versus placebo surgery;
- arthroscopic surgery versus exercise;
- arthroscopic surgery versus glucocorticoid injections;
- arthroscopic surgery versus non-arthroscopic lavage;
- arthroscopic surgery versus anti-inflammatory drugs; and
- arthroscopic surgery versus hyaluronic acid injections.

We pooled outcomes that assessed the harms of treatment (serious and total adverse events, progression of knee osteoarthritis, knee surgery (replacement or osteotomy)) across all studies (arthroscopic surgery versus any control).

Expecting some differences in the effect of the intervention across studies, we used a random-effects model as the default.

Subgroup analysis and investigation of heterogeneity

We performed subgroup analyses to assess if there were differences in pain and function for the primary comparison of arthroscopic surgery versus placebo:

- between participants with and without meniscal tear;
- between studies that describe arthroscopy with supervised exercise compared to studies that do not (or are unsupervised).

We performed subgroup analyses only for studies of arthroscopic surgery compared with placebo. We used the formal test for subgroup interactions in Review Manager (Review Manager 2020). We also compared the magnitude of the effects between the subgroups by means of assessing the overlap of the confidence intervals of the summary estimates. Non-overlap of the confidence intervals indicates statistical significance, and we used the formal test for differences in subgroups in Review Manager.

Sensitivity analysis

We performed sensitivity analyses to investigate the robustness of the effect on pain and function to potential selection and detection biases for the primary comparison of arthroscopic surgery versus placebo, at the primary time point (three months).

- We performed a sensitivity analysis with fixed-effect rather than random-effects model for pain and function for the primary comparison of arthroscopic surgery versus placebo, at the primary time point (three months).

- We also performed a sensitivity analysis by pooling outcomes that assessed benefit across all studies (arthroscopic surgery versus any control) as per our original review (Brignardello-Petersen 2017), for all time points.

We planned to investigate the robustness of the effect on pain and function to unit of analysis errors for the primary comparison of arthroscopic surgery versus placebo at the primary time point (three months), but no unit of analysis errors were identified.

Interpreting results and reaching conclusions

We followed guidance in Chapter 15 of the *Cochrane Handbookfor Systematic Reviews of Interventions* for interpreting results (Schünemann 2021a), and were aware of distinguishing a lack of evidence of effect from a lack of effect. We based our conclusions only on findings from the quantitative or narrative synthesis of included studies for this review. We avoided making recommendations for practice and our implications for research recommend priorities for future research and outline the remaining uncertainties in this area.

Summary of findings and assessment of the certainty of the evidence

We created a summary of findings table for arthroscopic surgery compared to placebo surgery using the following outcomes: pain, function, knee-specific health-related quality of life, treatment success, proportion experiencing serious and total adverse events, and proportion undergoing subsequent knee surgery (replacement or osteotomy). For benefits of surgery, we included pain, function and knee-specific quality of life measured at three months, and treatment success at last follow-up. For serious adverse events, total adverse events and knee replacements or osteotomies, we included combined data from all control groups at last follow-up.

Two review authors (DOC, RJ) independently assessed the certainty of the evidence across all studies contributing to the meta-analysis for each outcome, using the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias), as outlined in the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2021a; Schünemann 2021b). We developed the summary of findings table using GRADEpro software (GRADEPro GDT). We justified decisions to downgrade the quality of studies in the footnotes of the table.

We reported the number needed to treat for an additional beneficial outcome (NNTB) or the number needed to treat for an additional harmful outcome (NNTH), absolute and relative percent change in the 'Comments' column of the summary of findings table, as described in the Measures of treatment effect section above.

RESULTS

Description of studies

Results of the search

Overall, there are 16 included studies, 15 excluded studies, two studies awaiting classification and four ongoing studies.

The results of the search are shown in Figure 1. The search identified 3404 records (3044 from electronic databases, 359 from trial registries, 1 from reference checking). After removal of duplicates, we screened 2262 records. We retrieved 37 studies for full-text screening.

Figure 1. Study flow diagram

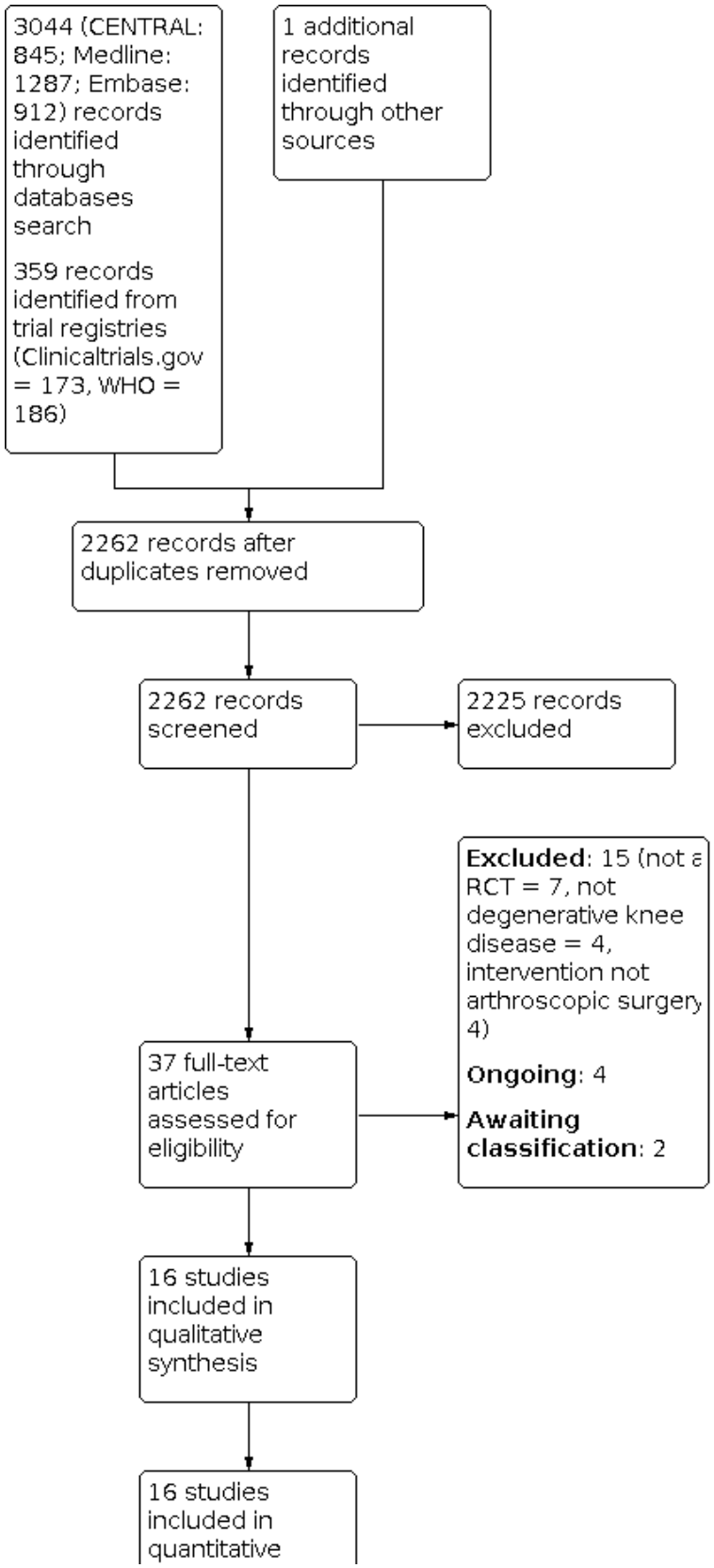
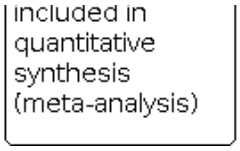


Figure 1. (Continued)



We excluded 15 studies (seven were not RCTs, four examined interventions other than arthroscopic surgery and four did not include participants with degenerative knee disease).

Sixteen trials met our criteria for inclusion (Chang 1993; Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Kise 2016; Merchan 1993; Moseley 1996; Moseley 2002; Osteras 2012; Roos 2018; Saeed 2015; Sihvonen 2013; Van de Graaf 2018; Vermesan 2013; Yim 2013).

Our previous review included 12 of these trials (i.e. Brignardello-Petersen 2017 included: Chang 1993; Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Kise 2016; Moseley 2002; Osteras 2012; Saeed 2015; Sihvonen 2013; Vermesan 2013; Yim 2013). Two other trials were published after our previous search cut-off date (Roos 2018; Van de Graaf 2018). In this update, we included two trials that were excluded from our previous review: Moseley 1996 was previously excluded as it included fewer than ten participants, while Merchan 1993 was previously erroneously excluded. Stensrud 2015 was included in our original review but was an interim preliminary report of a subset of 82 out of 140 participants from Kise 2016 (same clinical trial registry number NCT01002794).

We identified four ongoing trials in clinical trials registries (NCT02113280; NCT02995551; NCT04313569; NCT04837456).

We also note that an included study, Sihvonen 2013, has a separate trial registration for an ongoing 10 year follow-up (NCT01052233), but includes the same participants enrolled earlier. Thus, we have not counted this as a separate ongoing study, but grouped it as a secondary report of Sihvonen 2013.

Included studies

We provide a full description of the 16 included trials in the Characteristics of included studies tables. We contacted the authors of eight trials to retrieve (1) information about study design, participants, interventions and outcomes of the trial, (2) information required to complete the risk of bias assessments or (3) missing data for unreported or partially reported outcomes (Chang 1993; Herrlin 2007; Kise 2016; Moseley 1996; Roos 2018; Saeed 2015; Sihvonen 2013; Yim 2013). We received replies from the authors of four trials (Herrlin 2007; Kise 2016; Roos 2018; Sihvonen 2013).

Studies awaiting classification

Two studies are awaiting classification (Kang 2005; NCT00562822).

Ongoing studies

We identified four ongoing studies that did not have study results available at the time of submission of this review (NCT02113280; NCT02995551; NCT04313569; NCT04837456). One study compares arthroscopic meniscectomy to conservative treatment in people with degenerative meniscal tears (NCT04313569). The other trials are comparing knee arthroscopy to exercise (NCT02113280; NCT02995551; NCT04837456). We provide a description of these trials in the Characteristics of ongoing studies table.

Study design and setting

All 16 included trials were parallel-group RCTs. Fourteen trials included two intervention arms and two trials included three intervention arms (Moseley 1996; Moseley 2002).

The trials were conducted in Canada (Kirkley 2008), the USA (Chang 1993; Katz 2013; Moseley 1996; Moseley 2002), Denmark (Roos 2018), Finland (Sihvonen 2013), Sweden (Gauffin 2014; Herrlin 2007), Norway (Kise 2016; Osteras 2012), the Netherlands (Van de Graaf 2018), Spain (Merchan 1993), Italy (Vermesan 2013), Pakistan (Saeed 2015), and South Korea (Yim 2013).

Participant and intervention characteristics

Participant characteristics are detailed in the Characteristics of included studies tables, and age, osteoarthritis and meniscal tear criteria are shown in Table 1. A total of 2105 participants were included in the 16 trials. The number of participants per trial ranged from 10 in Moseley 1996 to 351 in Katz 2013. The minimum age requirements were 35 years in Roos 2018 to under 70 years in the Moseley 1996 and Moseley 2002 studies. The mean age of participants ranged from 46.4 years in Roos 2018 to 65 years in Chang 1993 (age reported for all trials except Moseley 1996 and Saeed 2015). Fifty-six percent of participants were female (gender reported for all trials except Moseley 1996). The mean duration of symptoms, reported in nine trials, ranged from 1.6 months in Osteras 2012 to 53 months in Chang 1993.

The required duration of knee pain varied between trials. Two trials specified persistent knee pain for more than two months (Kise 2016; Roos 2018); seven trials included participants with pain for more than three months (Chang 1993; Gauffin 2014; Herrlin 2007; Moseley 1996; Moseley 2002; Osteras 2012; Sihvonen 2013); and one trial specified symptoms of torn menisci for at least four weeks (Katz 2013).

The inclusion criteria regarding the presence/absence of osteoarthritis and degenerative meniscal tears varied across trials (Table 1). We describe these below, grouped according to the comparator to arthroscopic surgery.

Detailed descriptions of the interventions delivered in each trial are summarised in the Characteristics of included studies tables. We present a summary of the arthroscopic technique and comparison in each trial in Table 2. Arthroscopic procedures varied from debridement of torn menisci to surgical resection of proliferative synovium and excision of loose cartilage fragments. We describe the interventions below, grouped according to the comparator to arthroscopic surgery.

Arthroscopic surgery versus placebo surgery

Four trials compared arthroscopic surgery to placebo (Moseley 1996; Moseley 2002; Roos 2018; Sihvonen 2013). Two trials specifically included people with knee pain and a medial meniscal tear on MRI with minimal radiographic evidence of osteoarthritis

(either Kellgren-Lawrence (KL) classification grade 0 to 1 (Sihvonen 2013), or KL grade 0 to 2 (Roos 2018)). The other two trials included people with American College of Rheumatology (ACR)-defined or clinically diagnosed knee osteoarthritis and did not make any distinction based upon the presence/absence of degenerate meniscal tears (Moseley 1996; Moseley 2002).

All four placebo-controlled trials debrided degenerative or torn menisci if present. Two performed partial medial meniscectomy alone (Roos 2018; Sihvonen 2013), while the other two trials debrided degenerate articular cartilage as well as any torn or degenerate meniscal fragments (Moseley 1996; Moseley 2002). The placebo surgery control was similar across all four trials (see Table 2). Two trials included an additional arthroscopic lavage control arm (Moseley 1996; Moseley 2002), but these data were not extracted for this review.

Arthroscopic surgery versus exercise

Five trials compared arthroscopic surgery and exercise to exercise alone (Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Yim 2013), while three trials compared arthroscopic surgery to exercise therapy (Kise 2016; Osteras 2012; Van de Graaf 2018). We have considered these eight trials together.

One trial was performed primarily to investigate the value of arthroscopic treatment versus exercise for osteoarthritis (KL grades 2 to 4 excluding grade 4 if involving both compartments) (Kirkley 2008). This trial excluded people with large meniscal tears (bucket handle tears) detected by clinical examination or MRI in a minority of cases.

The remaining seven trials included participants with meniscal tears and excluded people with severe osteoarthritis (KL grade 4 (Katz 2013; Van de Graaf 2018), KL grade 3 or more (Kise 2016; Osteras 2012), KL grade 2 or more (Yim 2013), Ahlbacks classification grade 2 or more (Herrlin 2007), and 50% or more joint space narrowing (Gauffin 2014). Three trials specified a medial meniscal tear (Herrlin 2007; Kise 2016; Yim 2013), while the other four trials did not specify the site of the meniscal tear (Gauffin 2014; Katz 2013; Osteras 2012; Van de Graaf 2018). Gauffin 2014 included people with suspected meniscal injury but not all participants were found to have meniscal tears (see Table 1).

One trial performed a meniscal resection only if participants were found to have a meniscal tear at arthroscopy (Gauffin 2014). All other trials performed a partial meniscectomy (Herrlin 2007; Katz 2013, Kirkley 2008; Kise 2016; Osteras 2012; Van de Graaf 2018; Yim 2013). Three trials reported that they also performed limited debridement of articular cartilage if needed (Katz 2013; Kirkley 2008; Yim 2013).

Exercises varied across the trials, in both the control arm and in the post-operative exercise prescribed in the arthroscopy groups (see Table 2). Co-interventions were applied equally across both treatment groups.

Arthroscopic surgery versus other interventions

Vermesan 2013 included participants with medial compartment cartilage and meniscus lesions in MRI in a trial that compared arthroscopic surgery to a single intra-articular glucocorticoid injection.

Chang 1993 included participants with osteoarthritis KL grades 1 to 3 and did not specify criteria regarding the menisci in a trial that compared arthroscopic surgery to non-arthroscopic lavage.

Merchan 1993 included participants with mild osteoarthritis only (minimal joint space narrowing and formation of small osteophytes and did not specify criteria regarding the menisci) in a trial that compared arthroscopic surgery to non-steroidal anti-inflammatory drugs (NSAIDs).

Saeed 2015 included participants with osteoarthritis KL grade 2 and 3 only and did not specify criteria regarding presence/absence of meniscal tears in a trial that compared arthroscopic surgery to five intra-articular hyaluronic acid injections given at weekly intervals.

No trials compared arthroscopic surgery to supplements or complementary therapies, or both.

Outcomes

The pain, function, health-related quality of life and participant-reported success outcomes that were extracted for the purpose of analyses are summarised in Table 3.

All but one trial measured pain (Merchan 1993), although a second trial measured pain as part of the Oxford Knee Score but did not report the pain subscale result separately (Vermesan 2013). One trial used the WOMAC pain subscale to measure pain (Kirkley 2008), while five trials used the KOOS pain subscale (KOOS-P) (Gauffin 2014; Herrlin 2007; Katz 2013; Kise 2016; Roos 2018). Osteras 2012 measured pain using KOOS but did not report the pain subscale result separately.

Four trials measured pain using a Visual Analogue Scale (VAS) (Herrlin 2007; Osteras 2012; Van de Graaf 2018; Yim 2013), and two used a numerical rating scale (Moseley 1996; Sihvonen 2013). Two trials used the pain subscale of the Arthritis Impact Measurement Scale (AIMS2-P) (Chang 1993; Moseley 2002), one used the Knee Society Score System (KSSS) (Saeed 2015), and one used the Knee-Specific Pain Scale (KSPS) and the pain subscale of the SF-36 (Moseley 2002).

All but one trial measured function (Saeed 2015). Three trials used the composite KOOS score (Kise 2016; Osteras 2012; Roos 2018), and four used the KOOS ADL subscale (Gauffin 2014; Herrlin 2007; Kise 2016; Roos 2018) (to note: Osteras 2012 did not report the KOOS ADL subscale result separately). Three trials used the Lysholm Knee Score (Herrlin 2007; Sihvonen 2013; Yim 2013), and two used the WOMAC physical function subscale (Katz 2013; Kirkley 2008). Other measures of function included the WOMAC total score (Kirkley 2008), the Physical Functioning Scale (Moseley 2002), the Oxford Knee Score (Vermesan 2013), the Subjective Knee Form of the International Knee Documentation Committee (IKDC) (Van de Graaf 2018), the modified Hospital for Special Surgery Knee Rating Scale (mHSSKRS) (Merchan 1993), the McMaster-Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) (Kirkley 2008), the Arthritis Self-Efficacy Scale (ASES) (Kirkley 2008), AIMS physical function subscale (Chang 1993), the AIMS2 walking-bending subscale (Moseley 2002), the Tegner Activity Scale (Herrlin 2007; Yim 2013; Van de Graaf 2018), SF-36 physical function subscale (Moseley 2002; Katz 2013), SF-36 Physical Component Summary (Kirkley 2008; Kise 2016; Roos 2018; Van de Graaf 2018), a mobility (activity) scale developed for the trial (Moseley 1996),

and other physical performance tests (e.g. hop test, bend test) (Kise 2016; Roos 2018; Van de Graaf 2018).

Five trials measured knee-specific health-related quality of life (Gauffin 2014; Herrlin 2007; Kise 2016; Roos 2018; Sihvonen 2013). Four used the KOOS quality of life (QoL) subscale (Gauffin 2014; Herrlin 2007; Kise 2016; Roos 2018), and one used the Western Ontario Meniscal Evaluation Tool (WOMET) QoL score (Sihvonen 2013).

Seven trials measured generic health-related quality of life (Gauffin 2014; Katz 2013; Kirkley 2008; Kise 2016; Moseley 1996; Roos 2018; Sihvonen 2013). Two used the SF-36 Mental Component Summary (Kise 2016; Roos 2018), and four used the EQ-5D (Gauffin 2014; Katz 2013; Roos 2018; Van de Graaf 2018). Van de Graaf 2018 reported in their protocol that they measured the SF-36 but did not report the Mental Component Summary results. Other measures included the EuroQol Visual Analogue Scale (EQ VAS) (Gauffin 2014), the 15D (Sihvonen 2013), the standard-gamble utility technique (Kirkley 2008), and a general well-being scale developed for the trial (details of measurement scale not reported) (Moseley 1996).

Eight trials included participant-reported treatment success (Chang 1993; Gauffin 2014; Katz 2013; Merchan 1993; Moseley 1996; Roos 2018; Sihvonen 2013; Yim 2013). Various ways of measuring treatment success were used, including: improvement in the KOOS pain score of more than 10 points (Gauffin 2014); improvement in the WOMAC physical function score of at least 8 points (Katz 2013); improvement in the modified Hospital for Special Surgery Knee Rating Score of at least 10 points (Merchan 1993); improvement of at least 2 points on a 7-point Global Perceived Effect scale (Roos 2018); reduction of 1 cm or more in baseline overall well-being rated on a 10 cm visual analogue scale (0 = best to 10 = worst) (Chang 1993); participants reporting being 'much better' or 'better' on 5-point Likert scale in response to the question 'Is your knee better than before the intervention?' (Sihvonen 2013); participants reporting being 'very satisfied' or 'satisfied' with treatment (Yim 2013); participants reporting 'strongly agree' or 'slightly agree' to the question 'Do you feel the operation was worthwhile?' (Moseley 1996); and participant rating on a 6-point Likert scale ranging from 'delighted' to 'terrible' (Gauffin 2014).

Nine trials reported adverse events (Gauffin 2014; Katz 2013; Kise 2016; Merchan 1993; Moseley 2002; Roos 2018; Saeed 2015; Sihvonen 2013; Van de Graaf 2018). Of these, eight studies reported non-serious (other) adverse events (Gauffin 2014; Katz 2013; Kise 2016; Moseley 2002; Saeed 2015; Merchan 1993; Roos 2018; Van de Graaf 2018). Five of the nine studies reported serious adverse events, defined as an event that leads to hospitalisation, disability or death (Katz 2013; Merchan 1993; Roos 2018; Sihvonen 2013; Van de Graaf 2018). Seven studies did not report adverse events (Chang 1993; Herrlin 2007; Kirkley 2008; Moseley 1996; Osteras 2012; Vermesan 2013; Yim 2013). Moseley 2002 did not report adverse events details per group.

Six trials reported progression of knee osteoarthritis (Gauffin 2014; Herrlin 2007; Kise 2016; Sihvonen 2013; Van de Graaf 2018; Yim 2013), reporting it as a dichotomous outcome in all studies except Van de Graaf 2018. In three trials (Gauffin 2014; Kise 2016; Yim 2013), radiographic osteoarthritis was defined as equal to or higher than Kellgren-Lawrence (KL) grade 2 (definite osteophytes and possible joint space narrowing) (Kellgren 1957). Herrlin 2007 defined progression of osteoarthritis post hoc as at least one grade worse on the Ahlback classification (Ahlback 1968). Sihvonen 2013 defined progression as at least one grade progression in radiographic tibiofemoral knee osteoarthritis using the Kellgren-Lawrence classification. Van de Graaf 2018 reported mean Kellgren-Lawrence classification scores and not number of participants with osteoarthritis progression.

Four trials reported the proportion of participants undergoing subsequent knee surgery (replacement or osteotomy) (Katz 2013; Kise 2016; Sihvonen 2013; Van de Graaf 2018).

We contacted the authors of trials to retrieve missing data where outcome data were not fully reported. We received missing outcome data from the authors of four trials (Herrlin 2007; Kise 2016; Roos 2018; Sihvonen 2013).

Trial funding

Of the nine trials reporting their funding source (Chang 1993; Katz 2013; Kirkley 2008; Kise 2016; Merchan 1993; Moseley 2002; Roos 2018; Sihvonen 2013; Van de Graaf 2018), none received funding from industry. The other seven trials did not report any funding source (Gauffin 2014; Herrlin 2007; Moseley 1996; Osteras 2012; Saeed 2015; Vermesan 2013; Yim 2013).

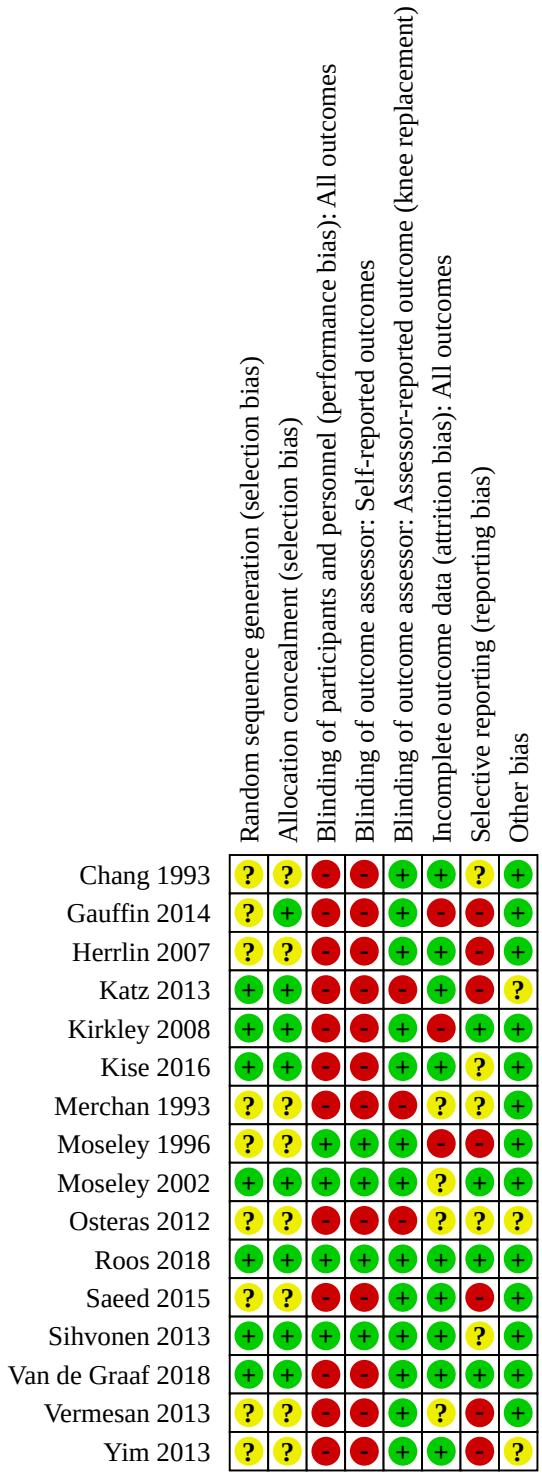
Excluded studies

A full description of all excluded trials is provided in the Characteristics of excluded studies table. As noted above, we excluded 15 studies (seven were not RCTs, four examined interventions other than arthroscopic surgery and four did not include participants with degenerative knee disease).

Risk of bias in included studies

All trials were susceptible to bias. Overall, 9 out of 16 trials (56%) were at risk of selection bias, 12 of 16 (75%) were susceptible to performance bias, 12 (75%) were at risk of detection bias; 7 (44%) were at risk of attrition bias, and 12 (75%) at risk of selective reporting bias (see Figure 2).

Figure 2. Risk of bias summary: review authors' judgements about each risk of bias item for each included study



Allocation

Seven out of 16 trials (44%) reported using adequate methods of randomisation and allocation concealment and were judged to be at low risk of selection bias (Katz 2013; Kirkley 2008; Kise 2016; Moseley 2002; Roos 2018; Sihvonen 2013; Van de Graaf 2018). Eight trials did not report the method of sequence generation (Gauffin 2014; Herrlin 2007; Merchan 1993; Moseley 1996; Osteras 2012; Saeed 2015; Vermesan 2013; Yim 2013), and seven trials did not clearly describe methods for concealing the allocation sequence (Herrlin 2007; Merchan 1993; Moseley 1996; Osteras 2012; Saeed 2015; Vermesan 2013; Yim 2013). Therefore, we judged the risk of selection bias as unclear in these trials. Additionally, baseline differences in outcome measures between the treatment groups were found in one trial (Merchan 1993).

Chang 1993 redesigned their study to a 'pre-randomised' design due to poor recruitment. In this method, also called the Zelen method of randomisation, participants were asked if they would be willing to undergo arthroscopic surgery, and those who indicated they would be, were allocated to either arthroscopic surgery or conservative treatment, then asked to consent to their allocated treatment. The authors did not report if the randomisation sequence was concealed.

Blinding

Only 4 of 16 (25%) trials were judged at low risk of performance bias, where both participants and study personnel were blinded to group assignment (Moseley 1996; Moseley 2002; Roos 2018; Sihvonen 2013). We judged the remaining 12 trials (75%) to be at high risk of performance bias as participants and study personnel were not blinded to group assignment (Chang 1993; Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Kise 2016; Osteras 2012; Van de Graaf 2018; Yim 2013), or blinding was not reported and probably was not done (Merchan 1993; Saeed 2015; Vermesan 2013).

All 16 trials reported using one or more self-reported outcomes. Four trials (25%) were at low risk of detection bias because participants were blinded to group allocation (Moseley 1996; Moseley 2002; Roos 2018; Sihvonen 2013). We judged the remaining 12 trials (75%) to be at high risk of detection bias, either because the studies reported that participants were not blinded to group allocation or blinding was not reported and probably was not done (Chang 1993; Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Kise 2016; Merchan 1993; Osteras 2012; Saeed 2015; Van de Graaf 2018; Vermesan 2013; Yim 2013).

Assessor-rated outcomes of interest (e.g. knee replacement, progression of knee osteoarthritis, adverse events) were measured in 10 of 16 (63%) trials (Gauffin 2014; Herrlin 2007; Katz 2013; Kise 2016; Merchan 1993; Roos 2018; Saeed 2015; Sihvonen 2013; Van de Graaf 2018; Yim 2013). There was a low risk of detection bias for assessor-rated outcomes in 5 of these 10 trials (50%), as assessors were effectively blinded to the treatment assignment (Kise 2016; Sihvonen 2013; Roos 2018; Van de Graaf 2018; Yim 2013). We judged a further three trials (30%) to be at low risk of detection bias despite no or unclear blinding of outcome assessment, as the outcomes were unlikely to be influenced by a lack of blinding (Gauffin 2014; Herrlin 2007; Saeed 2015). We assessed two trials (20%) as having a high risk of detection bias as outcome assessors were not adequately blinded (Katz 2013; Merchan 1993).

Incomplete outcome data

Nine out of 16 trials (56%) were judged at low risk of attrition bias because they had no withdrawals or losses to follow-up or the attrition was so small it was unlikely to have biased the results (Chang 1993; Herrlin 2007; Katz 2013; Kise 2016; Roos 2018; Saeed 2015; Sihvonen 2013; Van de Graaf 2018; Yim 2013). In three trials (19%), there was differential dropout across groups, we judged these studies to be at high risk of attrition bias (Gauffin 2014; Kirkley 2008; Moseley 1996). In the remaining four trials (25%), the reasons for incomplete outcome data were not reported, and we judged the risk of attrition bias as unclear (Merchan 1993; Moseley 2002; Osteras 2012; Vermesan 2013).

Selective reporting

We assessed 4 out of 16 trials (25%) to be at low risk of reporting bias (Kirkley 2008; Moseley 2002; Roos 2018; Van de Graaf 2018). Seven trials (44%) were judged at high risk of reporting bias (Gauffin 2014; Herrlin 2007; Katz 2013; Moseley 1996; Saeed 2015; Vermesan 2013; Yim 2013), while the remaining five trials (31%) were judged as having an unclear risk (Chang 1993; Kise 2016; Merchan 1993; Osteras 2012; Sihvonen 2013). Nine trials (56%) were not prospectively registered (Chang 1993; Herrlin 2007; Merchan 1993; Moseley 1996; Moseley 2002; Osteras 2012; Saeed 2015; Vermesan 2013; Yim 2013), and seven trials (44%) did not report one or more pre-specified outcomes or did not report outcomes as described in methods (Gauffin 2014; Herrlin 2007; Katz 2013; Kise 2016; Moseley 1996; Saeed 2015; Vermesan 2013). Two trials (13%) had incomplete reporting of outcome data (i.e. no standard deviations or confidence intervals) (Moseley 1996; Yim 2013).

Other potential sources of bias

Thirteen trials (81%) had no apparent other sources of bias (Chang 1993; Gauffin 2014; Herrlin 2007; Kirkley 2008; Kise 2016; Merchan 1993; Moseley 1996; Moseley 2002; Roos 2018; Saeed 2015; Sihvonen 2013; Van de Graaf 2018; Vermesan 2013). We judged the remaining three trials as having an unclear risk of other bias. In Katz 2013, 30.2% of participants assigned to physical therapy crossed over to arthroscopy within six months of randomisation, potentially underestimating any effect of surgery. It was unclear whether or not Osteras 2012 performed an unplanned interim analysis. In Yim 2013, an unspecified number of participants in the arthroscopic surgery group were not prescribed exercise. We did not identify any unit of analysis errors.

Effects of interventions

See: **Summary of findings 1** Summary of findings

See Summary of findings 1 for the main comparison.

Benefits

Arthroscopic surgery versus placebo surgery

Four trials compared arthroscopic surgery to placebo (Moseley 1996; Moseley 2002; Roos 2018; Sihvonen 2013).

Pain

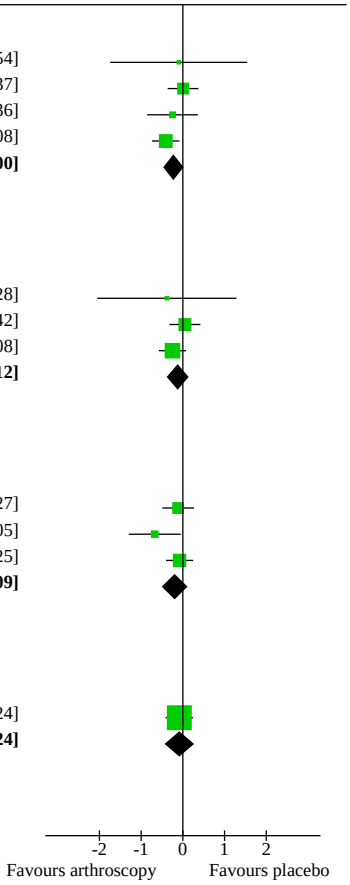
All four trials reported pain at up to three months. Three trials reported pain between three and six months (Moseley 1996; Moseley 2002; Sihvonen 2013), and three trials reported pain between six months and two years) (Moseley 2002; Roos 2018;

Sihvonen 2013). Sihvonen 2013 reported pain at five years. No trials reported pain between 5 and 10 years. Moseley 1996 used a 10-point numerical rating scale (lower score = less pain) to measure pain. Moseley 2002 used three pain scales: the Knee-Specific Pain Scale (KSPS) (0 to 100 scale, lower score = less pain; included in this review), the pain subscale of the Arthritis Impact Measurement Scales (AIMS2-P) (0 to 100 scale, lower score = less pain) and a 2-item pain subscale of the SF-36 (SF-36-P) (0 to 100 scale, higher score = less pain). Roos 2018 used the pain subscale of the Knee Injury and Osteoarthritis Outcome Score (KOOS-P) (0 to 100 scale, higher score = less pain). Sihvonen 2013 used an 11-point numerical rating scale (lower score = less pain).

High-certainty evidence indicates arthroscopic surgery leads to little or no difference in pain at up to three months compared with placebo surgery (standardised mean difference (SMD) -0.23, 95% confidence interval (CI) -0.45 to -0.001; I^2 = 0%; 4 trials, 309 participants; high-certainty evidence; Analysis 1.1; Figure 3). Mean post-operative pain in the placebo group at up to three months was 40.1 points on a 0 to 100 scale (where lower score indicates less pain) compared to 35.5 points in the arthroscopic surgery group, a difference of 4.6 points (95% CI 0.02 better to 9 better). This is an absolute improvement of 5% (95% CI 0.02% better to 9% better) and relative improvement of 8% (95% CI 0.03% better to 15% better). As the 95% confidence intervals do not include any appreciable benefit, we did not downgrade for imprecision.

Figure 3. Forest plot of comparison 1: arthroscopic surgery versus placebo surgery, outcome: 1.1 Pain (lower score = less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo surgery Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.1.1 Up to 3 months								
Moseley 1996	4.5	2.13	2	4.8	2.67	5	1.9%	-0.10 [-1.74 , 1.54]
Moseley 2002	-46.8	21.9	58	-46.9	24.9	56	37.5%	0.00 [-0.36 , 0.37]
Roos 2018	-71.9	17.6	21	-67	21	21	13.7%	-0.25 [-0.86 , 0.36]
Sihvonen 2013	3.1	2.13	70	4.1	2.67	76	46.9%	-0.41 [-0.74 , -0.08]
Subtotal (95% CI)			**151**			**158**	**100.0%**	**-0.23 [-0.45 , -0.00]**
Heterogeneity: Tau² = 0.00; Chi² = 2.75, df = 3 (P = 0.43); I² = 0%								
Test for overall effect: Z = 1.98 (P = 0.05)								
1.1.2 >3 months up to 6 months								
Moseley 1996	4.5	2.35	2	5.6	2.43	5	2.1%	-0.38 [-2.05 , 1.28]
Moseley 2002	-45.1	20.6	55	-46.3	26.4	57	42.7%	0.05 [-0.32 , 0.42]
Sihvonen 2013	2.5	2.35	70	3.1	2.43	76	55.2%	-0.25 [-0.58 , 0.08]
Subtotal (95% CI)			**127**			**138**	**100.0%**	**-0.12 [-0.37 , 0.12]**
Heterogeneity: Tau² = 0.00; Chi² = 1.51, df = 2 (P = 0.47); I² = 0%								
Test for overall effect: Z = 1.01 (P = 0.31)								
1.1.3 >6 months up to 2 years								
Moseley 2002	-45	23	52	-42.3	24.2	55	37.3%	-0.11 [-0.49 , 0.27]
Roos 2018	-79.1	17.2	22	-63.1	28.6	20	17.7%	-0.67 [-1.30 , -0.05]
Sihvonen 2013	2.7	2.54	70	2.9	2.45	76	45.1%	-0.08 [-0.40 , 0.25]
Subtotal (95% CI)			**144**			**151**	**100.0%**	**-0.20 [-0.48 , 0.09]**
Heterogeneity: Tau² = 0.02; Chi² = 2.88, df = 2 (P = 0.24); I² = 30%								
Test for overall effect: Z = 1.35 (P = 0.18)								
1.1.4 >2 years up to 5 years								
Sihvonen 2013	2	2.3	68	2.2	2.4	74	100.0%	-0.08 [-0.41 , 0.24]
Subtotal (95% CI)			**68**			**74**	**100.0%**	**-0.08 [-0.41 , 0.24]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.50 (P = 0.61)								



There was little to no difference in pain with arthroscopic surgery compared with placebo: between three and six months (SMD -0.12, 95% CI -0.37 to 0.12; or 2.4 points better (7.4 points better to 2.4 points worse) on a 0 to 100 point scale; I^2 = 0%; 3 trials, 265 participants); between six months and two years (SMD -0.20, 95% CI -0.48 to 0.09; or equivalent to 4 points better (9.6 points better to 1.8 points worse) on a 100-point scale; I^2 = 30%; 3 trials, 295 participants); or between two and five years (last follow-up) (SMD -0.08, 95% CI -0.41 to 0.24; or 1.6 points better (8.2 points better to 4.8 points worse) on a 0 to 100 points scale; 1 study, 142 participants).

Subgroup analyses based on the presence or absence of meniscal tear and use or not of supervised exercise did not alter the estimates of treatment effect to a clinically important level (Analysis 8.1; Analysis 8.2; Analysis 8.3; Analysis 8.4; Analysis 9.1; Analysis 9.2; Analysis 9.3; Analysis 9.4).

Sensitivity analyses indicated there was little change in effect estimates when restricting studies to those at low risk of selection bias. As no studies were at risk of detection biases, no sensitivity analysis for this bias could be performed. Use of a fixed-effect model also did not alter the results.

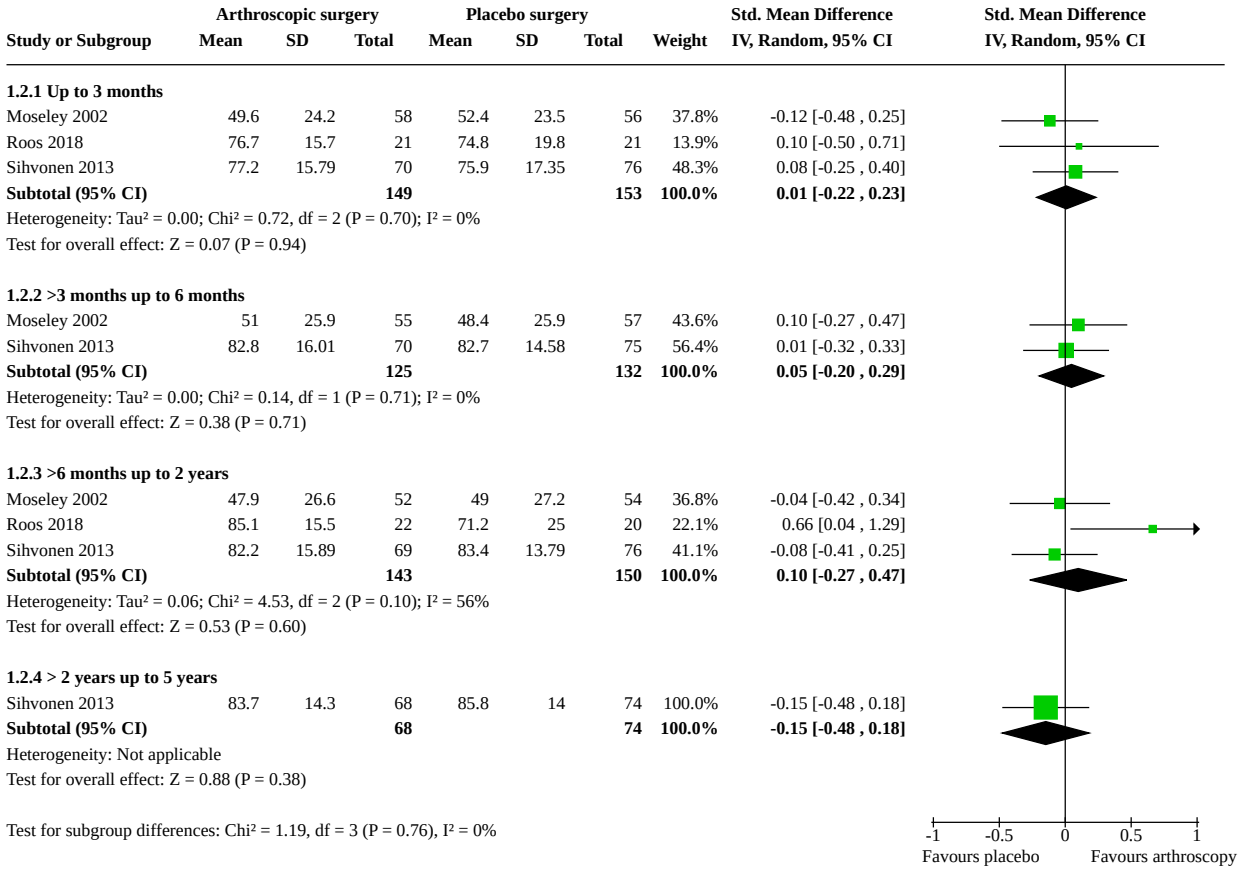
Sensitivity analysis comparing arthroscopic surgery to any control indicates arthroscopic surgery leads to little reduction in pain at up at three months (SMD -0.21, 95% CI -0.32 to -0.10; equivalent to 4.2 points better (95% CI 2 to 6.4 points better) on a 0 to 100 scale; I^2 = 0%; 12 trials, 1283 participants; high-certainty evidence), between three and six months (SMD -0.19, 95% CI -0.30 to -0.07; or 3.8 points better (1.4 to 6 points better) on a 0 to 100 scale); I^2 = 0%; 8 trials, 1252 participants; high-certainty evidence), between six months and two years (SMD -0.11, 95% CI -0.22 to -0.01; or 2.2 points better (0.2 to 4.4 points better) on a 100 point scale; I^2 = 0%; 11 trials, 1505 participants; high-certainty evidence), and leads to no difference in pain between two and five years (SMD -0.01, 95% CI -0.30 to 0.29; I^2 = 49%; 3 trials, 361 participants), compared with any control (Analysis 13.1).

Function

Three trials reported function at up to three months (Moseley 2002; Roos 2018; Sihvonen 2013). These same three trials reported function between six months and two years, using a variety of scales: the SF-36 Physical Function subscale and the Arthritis Impact Measurement Scales walking-bending subscale (AIMS2-WB) (Moseley 2002); the composite Knee Injury and Osteoarthritis Outcome Score (KOOS5), the KOOS ADL subscale and the SF-36 Physical Component Summary (Roos 2018); and Lysholm Knee Score (Sihvonen 2013). Two trials reported function between three and six months (Moseley 2002; Sihvonen 2013). Sihvonen 2013 reported function at five years. No trials reported function between 5 and 10 years.

High-certainty evidence indicates arthroscopic surgery leads to little or no difference in knee function at up to three months compared with placebo (SMD 0.01, 95% CI -0.22 to 0.23; I^2 = 0%; 3 trials, 302 participants; high-certainty evidence; Analysis 1.2; Figure 4). Mean post-operative function in the placebo group at up to three months was 75.9 points on a 0 to 100 rating scale (where higher score indicates better function) compared to 76 points in the arthroscopic surgery group, a difference of 0.1 points better (95% CI 3.2 worse to 3.4 better). This is an absolute improvement of 0.1% (95% CI 3% worse to 3% better) and relative improvement of 0.2% (95% CI 5% worse to 6% better). As the 95% confidence intervals do not include any appreciable benefit, we did not downgrade for imprecision. There was little or no difference in knee function with arthroscopic surgery compared with placebo: between three and six months (SMD 0.05, 95% CI -0.20 to 0.29; equivalent to 0.7 points better (2.9 worse to 4.2 better) on a 0 to 100 point scale; I^2 = 0%; 2 trials, 257 participants); between six months and two years (SMD 0.10, 95% CI -0.27 to 0.47; or 1.5 points better (3.9 points worse to 6.9 points better) on a 0 to 100 point scale; I^2 = 56%; 3 trials, 293 participants); or between two and five years (last follow-up) (SMD -0.15, 95% CI -0.48 to 0.18; equivalent to 2.2 points worse (7 points worse to 2.6 points better) on a 0 to 100 point scale; 1 trial, 142 participants).

Figure 4. Forest plot of comparison 1: arthroscopic surgery versus placebo surgery, outcome: 1.2 Function (higher score = better function)



Subgroup analyses based on the presence or absence of meniscal tear and use or no use of supervised exercise did not alter the estimates of treatment effect to a clinically important level (Analysis 8.5; Analysis 8.6; Analysis 8.7; Analysis 8.8; Analysis 9.5; Analysis 9.6; Analysis 9.7; Analysis 9.8).

Sensitivity analyses to investigate the robustness of the effects to potential selection and detection biases or use of a fixed-effect model did not alter the results. Sensitivity analysis comparing arthroscopic surgery to any control indicates arthroscopic surgery probably leads to a very little improvement in function at up to three months (SMD 0.19, 95% CI 0.04 to 0.34; 2.8 points better (0.6 to 5.0 better) on a 0 to 100 point scale; I^2 = 45%; 12 trials, 1403 participants; moderate-certainty evidence), and little or no difference between three and six months (SMD 0.10, 95% CI -0.01 to 0.21; or 1.5 points better (0.1 points worse to 3.1 better) on a 0 to 100 point scale; I^2 = 0%; 7 trials, 1245 participants; high-certainty evidence), between six months and two years (SMD 0.11, 95% CI 0.01 to 0.20; or 1.6 points better (0.1 to 2.9 points better) on a 0 to 100 point scale; I^2 = 0%; 12 trials, 1651 participants; high-certainty evidence), and between six months and two years (SMD -0.09, 95% CI -0.30 to 0.12; I^2 = 0%; 3 trials, 361 participants) compared with any control (Analysis 13.2).

Knee-specific health-related quality of life

Knee-specific health-related quality of life was reported in two trials at up to three months (Roos 2018; Sihvonen 2013), and between six months and two years, using the KOOS quality of life subscale and the WOMET score, respectively. The Sihvonen 2013 study also reported knee-related quality of life between three and six months, and between two and five years. Neither trial reported knee-related quality of life between 5 and 10 years.

Moderate-certainty evidence indicates arthroscopic surgery probably leads to little or no difference, or a very small improvement, in knee-specific quality of life at up to three months compared with placebo (SMD 0.31, 95% CI 0.02 to 0.59; translates to 5.6 points better (0.36 better to 10.7 points better) on a 0 to 100 point scale; I^2 = 0%; 2 trials, 188 participants; moderate-certainty evidence; Analysis 1.3). Mean post-operative knee-specific quality of life in the placebo group at up to three months was 69.7 points on a 0 to 100 rating scale (where higher score indicates better quality of life) compared with 75.3 points in the arthroscopic surgery group, a difference of 5.6 points better (95% CI 0.4 better to 10.7 better), an absolute improvement of 6% (0.4% better to 11% better) and a relative improvement of 11% (95% CI 0.8% better to 20% better). We downgraded the certainty of the evidence due to imprecision (the 95% confidence intervals include both an unimportant improvement and the clinically important improvement threshold of 10%). Moderate-certainty evidence also

indicates little or no difference at other time frames: between three and six months (MD 2.60, 95% CI -4.27 to 9.47; absolute improvement 3%, 95% CI 4% worse to 9% better; I^2 = 0%; 1 trial, 146 participants); between six months and two years (SMD 0.23, 95% CI -0.24 to 0.70; translates to 4.2 points better (4.3 points worse to 12.7 points better) on a 0 to 100 point scale; absolute improvement 6% (6% worse to 18% better); I^2 = 50%; 2 trials, 188 participants); or between two and five years (last follow-up) (SMD -0.02, 95% CI -0.35 to 0.31; equivalent to 0.4 points worse (6 points worse to 5.6 points better); 1 trial, 142 participants).

Sensitivity analysis comparing arthroscopic surgery to any control did not alter the estimates of treatment effect to a clinically important level (Analysis 13.3).

Generic health-related quality of life

Generic health-related quality of life was reported in one trial at 3 and 24 months (Roos 2018), using the SF-36 Mental Component Summary (MCS) and the EQ-5D. One additional trial reported generic quality of life using the 15D at 12 months of follow-up (Sihvonen 2013). No trials reported generic health-related quality of life between three and six months, or between two and 10 years.

Based on Roos 2018, arthroscopic surgery probably results in little or no difference in generic quality of life at up to three months compared with placebo (MD -3.50, 95% CI -7.20 to 0.20; 1 trial, 42 participants; moderate-certainty evidence; Analysis 1.4). The certainty of the evidence was downgraded due to serious imprecision (data from a single trial).

Moderate-certainty evidence from two trials - Roos 2018 and Sihvonen 2013 - indicates arthroscopic surgery probably results in little or no difference in generic quality of life between six months and two years compared with placebo (SMD 0.15, 95% CI -0.28 to 0.58; I^2 = 42%; 2 trials, 188 participants; moderate-certainty evidence). This translates to a mean difference in generic quality of life of 1.5 points (95% CI -2.8 to 5.8 points) on a 0 to 100 SF-36 scale (higher score = better quality of life) or 0.01 points (95% CI -0.2 to 0.03) on the 15D generic quality of life scale from 0 to 1 (higher score = better quality of life). We downgraded the certainty of the evidence due to serious indirectness (outcome dissimilarity).

Sensitivity analysis comparing arthroscopic surgery to any control did not alter the estimates of treatment effect to a clinically important level (Analysis 13.4).

Participant-reported treatment success

Three trials reported participant-reported treatment success at up to five years (Moseley 1996; Roos 2018; Sihvonen 2013). Moseley 1996 used the number of participants reporting they 'strongly agree' or 'slightly agree' for the question 'Do you feel the operation was worthwhile?' at three and six months. Roos 2018 used the number of participants rating their overall improvement in knee symptoms after the operation as 'better' or 'much better' at 3 and 24 months. Sihvonen 2013 used the number of participants reporting being 'much better' or 'better' for the question 'Is your knee better than before the intervention?' at 12, 24 and 60 months. No trials reported participant-reported treatment success between 5 and 10 years.

Low-certainty evidence indicates arthroscopic surgery may lead to little or no difference in participant-reported treatment success compared with placebo at up to five years (68 out of 91 participants in the arthroscopic surgery group (75%) versus 72 out of 98 participants in the placebo group (74%), RR 1.11, 95% CI 0.66 to 1.86; I^2 = 53%; 3 trials, 189 participants; low-certainty evidence; Analysis 1.5). We downgraded the certainty of the evidence due to serious indirectness (some diversity in definition and timing of outcome measurement: reported at 6 months, 24 months and 5 years across trials) and serious imprecision (small number of events).

Sensitivity analysis comparing arthroscopic surgery to any control did not alter the estimate of treatment effect to a clinically important level (RR 1.24, 95% CI 0.96 to 1.60; I^2 = 83%; 8 trials, 851 participants; Analysis 13.5).

Arthroscopic surgery versus exercise

Five trials compared arthroscopic surgery plus exercise therapy to exercise therapy alone (Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Yim 2013), and three trials compared arthroscopic surgery to exercise therapy (Kise 2016; Osteras 2012; Van de Graaf 2018). We have considered these eight trials together.

Pain

Seven trials reported pain at up to three months (Gauffin 2014; Herrlin 2007; Kirkley 2008; Kise 2016; Osteras 2012; Van de Graaf 2018; Yim 2013), five trials reported pain between three and six months (Herrlin 2007; Katz 2013; Kirkley 2008; Van de Graaf 2018; Yim 2013), seven trials between six months and two years (Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Kise 2016; Van de Graaf 2018; Yim 2013), and three trials between two and five years (Gauffin 2014; Katz 2013; Kise 2016). No trials reported pain between 5 and 10 years.

Moderate-certainty evidence indicates arthroscopic surgery leads to little or no benefit at any time point compared with exercise. At up to three months, the SMD was -0.21 (95% CI -0.33 to -0.08; I^2 = 0%; 7 trials, 942 participants). This translates to a small, clinically unimportant mean difference in pain of -4.2 points (-6.6 points to -1.6 points) on a 0 to 100 point pain scale, below the clinically important threshold of 15 points on 0 to 100 scale (or 15% absolute change). Between three and six months, the SMD was -0.20 (95% CI -0.33 to -0.08; I^2 = 0%; 5 trials, 987 participants; back-translated MD -4.0 points (-6.6 to -1.6), on a 0 to 100 scale). Between six months and two years, the SMD was -0.11 (95% CI -0.22 to 0.01; I^2 = 0%; 7 trials, 1178 participants; which translates to a mean difference of -2.2 points (-4.4 to 0.2 points) on a 0 to 100 scale). In the longer term (i.e. between two and five years), arthroscopic surgery may lead to little or no difference in pain compared with exercise (MD 1.27 points, 95% CI -8.50 to 11.03; I^2 = 75%; 2 trials, 219 participants; Analysis 2.1). We downgraded the certainty of the evidence by one grade due to serious risk of bias (largely detection bias). Other biases probably did not result in an overestimate of effect. We were unable to include data on pain at five years from Katz 2013 as it was not reported in a useable way.

Function

Seven trials reported function at up to three months (Gauffin 2014; Herrlin 2007; Kirkley 2008; Kise 2016; Osteras 2012; Van de Graaf 2018; Yim 2013), five trials reported function between three and six months (Herrlin 2007; Katz 2013; Kirkley 2008; Van de Graaf 2018; Yim 2013), seven trials between six months and two

years (Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Kise 2016; Van de Graaf 2018; Yim 2013), and three trials between two and five years (Gauffin 2014; Katz 2013; Kise 2016). There was no important heterogeneity at any time point. No trials reported function between 5 and 10 years.

Moderate-certainty evidence indicates arthroscopic surgery may lead to little or no difference in function compared with exercise: at up to three months (SMD 0.13, 95% CI 0.00 to 0.26; I^2 = 0%; 7 trials, 949 participants; translates to 1.9 points on a 0 to 100 point scale (0.01 to 3.8 points)); between three and six months (SMD 0.12, 95% CI -0.01 to 0.24; I^2 = 0%; 5 trials, 988 participants; translates to 1.8 (-0.1 to 3.5) points on 0 to 100 point scale); between six months and two years (SMD 0.10, 95% CI -0.01 to 0.21; I^2 = 0%; 7 trials, 1227 participants; 1.5 (-0.1 to 3.1 points) on 0 to 100 scale); and between two and five years (MD -0.79 points, 95% CI -5.50 to 3.91; I^2 = 20%; 2 trials, 219 participants; Analysis 2.2). We downgraded the certainty of the evidence by one grade due to serious risk of bias (largely detection bias). Other biases probably did not result in an overestimate of effect. We were unable to include data on function at five years from Katz 2013 as it was not reported in a useable way.

Knee-specific health-related quality of life

Three trials reported knee-specific health-related quality of life on the KOOS quality of life (QoL) subscale at up to three months (Gauffin 2014; Herrlin 2007; Kise 2016), one trial reported this outcome between three and six months (Herrlin 2007), three trials between six months and two years (Gauffin 2014; Herrlin 2007; Kise 2016), and two trials between two and five years (Gauffin 2014; Kise 2016). There was no important heterogeneity at any time point. No trials reported knee-specific health-related quality of life between 5 and 10 years.

Low-certainty evidence indicates arthroscopic surgery may lead to little or no difference, or a very small improvement, in knee-specific quality of life compared with exercise at up to three months (MD 6.87, 95% CI 2.55 to 11.19; I^2 = 0%; 3 trials, 347 participants). We downgraded the certainty of the evidence due to serious risk of bias (largely detection bias) and serious imprecision (the 95% confidence intervals include both an unimportant improvement and the clinically important improvement threshold of 10%). There was little or no difference in knee-specific quality of life with arthroscopic surgery compared with exercise between three and six months (MD 0.49, 95% CI -8.28 to 9.26; 1 trial, 96 participants), between six months and two years (MD 4.47, 95% CI -1.33 to 10.28; I^2 = 32%; 3 trials, 348 participants), and between two and five years (MD 2.13, 95% CI -5.74 to 10.00; I^2 = 35%; 2 trials, 220 participants; Analysis 2.3).

Generic health-related quality of life

Two trials reported generic health-related quality of life at up to three months (Gauffin 2014; Kirkley 2008), one trial between three and six months (Kirkley 2008), three trials between six months and two years (Gauffin 2014; Kirkley 2008; Kise 2016), and one trial between two and five years (Gauffin 2014). These trials used different measurement scales: EQ-5D (Gauffin 2014), standard-gamble utility score (Kirkley 2008), and SF-36 MCS (Kise 2016). There was no important heterogeneity. None of the trials reported generic health-related quality of life between 5 and 10 years.

We downgraded the evidence to low certainty at each time point due to the potential for biased and imprecise estimates. Low-certainty evidence (downgraded due to serious risk of bias and imprecision) indicates arthroscopic surgery may lead to little or no improvement in generic quality of life compared with exercise: at up to three months (SMD 0.09, 95% CI -0.14 to 0.32; I^2 = 0%; 2 trials, 290 participants), translates to 0.01 points (-0.01 to 0.02) on the 15D 0 to 1 scale, and 0.9 points (-1.4 to 3.2) points on the 0 to 100 SF-36 scale); between three and six months (MD 0.03, 95% CI -0.04 to 0.10, 15D 0 to 1 scale; I^2 = 0%; 1 trial, 163 participants); between six months and two years (SMD 0.03, 95% CI -0.16 to 0.22; I^2 = 0%; 3 trials, 425 participants; translates to 0.001 points (-0.01 to 0.01) on 0 to 1 scale and 0.3 points (-1.6 to 2.2) points on 0 to 100 scale); and between two and five years (MD -0.05, 95% CI -0.12 to 0.02, EQ-5D 0 to 1 scale; 1 trial, 101 participants; Analysis 2.4).

Participant-reported treatment success

Three trials reported participant-reported treatment success at up to five years (Gauffin 2014; Katz 2013; Yim 2013). No trials reported participant-reported treatment success between 5 and 10 years.

There was considerable heterogeneity in results (I^2 = 88%; Chi^2 = 16.04; P < 0.001), due to a larger treatment effect in one study (Katz 2013), but the studies are consistent in the effect direction. It is uncertain whether arthroscopic surgery leads to participant-reported treatment success at up to five years compared with exercise because the certainty of the evidence is low (RR 1.17, 95% CI 0.86 to 1.59; Analysis 2.5). The certainty was downgraded due to serious risk of bias (all trials had high risk of performance and detection bias, and some concerns with selection bias). There was also serious statistical inconsistency (I^2 = 88%; Chi^2 = 16.04; P < 0.001), which could be explained by small and large treatment effects across only three studies, and possible indirectness (diversity in outcome measurement across trials).

Arthroscopic surgery versus glucocorticoid injection

One trial compared arthroscopic surgery to a single intra-articular glucocorticoid injection (Vermesan 2013).

Function

Vermesan 2013 reported function at one month and one year using the Oxford Knee Score. Mean function at one month in the glucocorticoid injection group was 39.9 (0 to 48 scale, higher is better) compared with 42.8 in the arthroscopic surgery group, a difference of 2.9 points (95% CI 1.64 to 4.16; 1 trial, 120 participants; low-certainty evidence; Analysis 3.1). There was no difference in function with arthroscopic surgery compared with a single intra-articular glucocorticoid injection at one year (MD 1.40, 95% CI -0.07 to 2.87; 1 trial, 98 participants; low-certainty evidence). Based on this study, arthroscopic surgery may slightly improve function compared with a single intra-articular glucocorticoid injection at one month but leads to little or no difference in function at one year. We downgraded the evidence due to serious risk of bias (some concerns with selection, performance and detection bias, and high risk of reporting bias) and serious imprecision (only one study).

Outcomes not measured

Pain, knee-specific and generic health-related quality of life, treatment success, serious and total adverse events and progression of knee osteoarthritis were not measured.

Arthroscopic surgery versus non-arthroscopic lavage

One trial compared arthroscopic surgery plus physical therapy and analgesia to non-arthroscopic joint lavage plus physical therapy and analgesia (Chang 1993).

Pain

Chang 1993 reported pain at 3 and 12 months using the Arthritis Impact Measurement Scales pain subscale (AIMS-P; 0 to 10 scale; lower score indicates less pain). Based on this study, arthroscopic surgery may lead to little or no difference in pain compared with non-arthroscopic lavage at 3 months (MD -0.40, 95% CI -1.66 to 0.86; 1 trial, 32 participants; low-certainty evidence) and 12 months (MD 0.30, 95% CI -1.15 to 1.75; 1 trial, 32 participants; low-certainty evidence; Analysis 4.1). We downgraded the evidence due to serious risk of bias (some concerns with selection and reporting bias, and high risk of performance and detection bias) and serious imprecision (only one study).

Function

Chang 1993 reported function at 3 and 12 months using the Arthritis Impact Measurement Scales physical function subscale (AIMS-PF; 0 to 10 scale; lower score is better). Based on this study, arthroscopic surgery may lead to little or no difference in function compared with non-arthroscopic lavage at 3 months (MD 0.50, 95% CI -0.25 to 1.25; 1 trial, 32 participants; low-certainty evidence) and 12 months (MD 0.30, 95% CI -0.50 to 1.10; 1 trial, 32 participants; low-certainty evidence; Analysis 4.2). We downgraded the evidence due to serious risk of bias (some concerns with selection and reporting bias, and high risk of performance and detection bias) and serious imprecision (only one study).

Participant-reported treatment success

Chang 1993 reported participant-reported treatment success, defined as 1 cm improvement or greater from baseline global assessment of overall well-being (measured on a 0 to 10 visual analogue scale; lower is better). Seven of 16 participants in the arthroscopic surgery group and seven of 12 participants in the non-arthroscopic lavage group reported treatment success according to this definition. Based on this study, arthroscopic surgery may result in little or no difference in participant-reported treatment success compared with non-arthroscopic lavage at 12 months (last follow-up) (RR 0.75, 95% CI 0.36 to 1.56; 1 trial, 28 participants; low-certainty evidence; Analysis 4.3). We downgraded the evidence due to serious risk of bias (some concerns with selection and reporting bias, and high risk of performance and detection bias) and serious imprecision (only one study).

Outcomes not measured

Knee-specific and generic health-related quality of life, progression of knee osteoarthritis and knee surgery (replacement or osteotomy) were not measured. It was unclear if serious and total adverse events were measured as they were not reported.

Arthroscopic surgery versus non-steroidal anti-inflammatory drugs

One trial compared arthroscopic surgery and physiotherapy to non-steroidal anti-inflammatory drugs (NSAIDs) and physiotherapy (Merchan 1993).

Function

Merchan 1993 measured function at an average of 25 months follow-up (range 12 to 36 months) using the modified Hospital for Special Surgery Knee Rating Score (0 to 100; higher score indicates better function). However, only group means were reported and our attempts to obtain missing data from authors were unsuccessful.

Participant-reported treatment success

Merchan 1993 defined treatment success as a 10-point or greater increase in the post-treatment modified Hospital for Special Surgery Knee Rating Score (0 to 100; higher is better; contains a subjective and objective subscale). Twenty-six of 35 participants in the arthroscopic surgery plus physiotherapy group and six of 38 participants in the NSAIDs plus physiotherapy group were improved at last follow-up according to this definition. Based on this study, arthroscopic surgery plus physiotherapy may improve participant-reported treatment success compared with anti-inflammatory drugs plus physiotherapy at last follow-up (RR 4.70, 95% CI 2.20 to 10.06; 1 trial, 73 participants; low-certainty evidence; Analysis 5.1). We downgraded the evidence due to serious risk of bias (some concerns with selection and reporting bias, and high risk of performance and detection bias) and serious imprecision (only one study, wide confidence intervals).

Outcomes not measured

Pain, knee-specific and generic health-related quality of life, progression of knee osteoarthritis and knee surgery (replacement or osteotomy) were not measured.

Arthroscopic surgery versus hyaluronic acid injection

One trial compared arthroscopic surgery to receipt of five hyaluronic acid injections given at weekly intervals (Saeed 2015).

Pain

Saeed 2015 reported pain at one, three and six months using the Knee Society Scoring System (KSSS pain score of 30 or higher; higher score = less pain). Based on this study, it is uncertain whether arthroscopic surgery reduces pain at three and six months compared to hyaluronic acid injections because the certainty of the evidence is very low. We downgraded the evidence due to serious risk of bias (some concerns with selection, performance and detection bias, and high risk of reporting bias), serious imprecision (wide confidence intervals) and serious indirectness (outcome dissimilarity).

Outcomes not measured

Function, knee-specific and generic quality of life, treatment success, progression of knee osteoarthritis and knee surgery (replacement or osteotomy) were not measured.

Harms

Knee arthroscopy versus all control groups

Serious adverse events

Serious adverse events, defined as those necessitating hospitalisation (including subsequent knee surgery), prolonged inpatient hospital care, or those that are life threatening or result in death or disability, were reported in nine trials. This includes two placebo-controlled trials (Roos 2018; Sihvonen 2013), five trials with exercise control (Gauffin 2014; Herrlin 2007; Katz 2013; Kise

2016; Van de Graaf 2018), one trial with NSAIDs control (Merchan 1993), and one trial with a single intra-articular glucocorticoid injection control (Vermesan 2013).

In total, serious adverse events were reported in 32 of 574 participants (5.6%) in the control groups and in 41 of 632 participants (6.5%) in the arthroscopy groups from eight trials. Events included repeat arthroscopy, pulmonary embolism, deep vein thrombosis, heart attack, death, knee surgery, post-operative knee infection and anterior cruciate ligament reconstruction. Based on these studies, the risk of serious adverse events may increase with arthroscopic surgery at up to five years compared with control (RR 1.35, 95% CI 0.64 to 2.83; I^2 = 47%; 8 trials, 1206 participants; low-certainty evidence; Analysis 7.1; Summary of findings 1). We downgraded the certainty of the evidence due to serious imprecision (small number of events) and possible reporting bias (incomplete reporting of outcomes across studies).

Roos 2018 reported that two of 22 (9%) participants from the arthroscopic surgery group had serious knee-related adverse events (one partial meniscectomy, one anterior cruciate ligament reconstruction). Two other serious adverse events were reported (abdominal surgery, malignant melanoma) but as these two events were likely unrelated to the intervention, we excluded the events from Analysis 7.1.

Sihvonen 2013 reported that eight of 76 (10.5%) participants from the placebo group had serious knee-related adverse events (one proximal tibial osteotomy, seven arthroscopic partial meniscectomies) and no other serious adverse events. Sihvonen 2013 reported that seven of 70 (10%) participants from the arthroscopic surgery group had a serious knee-related adverse event (three knee replacements, four arthroscopies) and that one (1.4%) participant from this group had a serious adverse events (one deep infection at four months), giving a total of eight of 70 (11.4%) participants with serious adverse events in this group.

Gauffin 2014 reported no serious adverse events in the exercise group, and that three of 66 (4.5%) participants in the arthroscopic surgery group had serious adverse events (two repeat arthroscopies, one death after three years).

Herrlin 2007 reported no serious adverse events with exercise. However, 13 of 49 (26.5%) participants crossed over from the exercise group and had an arthroscopic procedure on average 6.5 months following treatment, and three of 47 (6.4%) participants from the arthroscopy group underwent an additional arthroscopic procedure at between 13 and 40 months following their original surgery. No other serious adverse events were reported.

Katz 2013 reported that, at 12 months, three of 109 (2.8%) participants from the physical therapy group had a serious knee-related adverse event (subsequent knee replacement) and two of 109 (1.8%) participants from this group had other serious adverse events (one stroke; one sudden death), giving a total of five of 109 (4.6%) participants with serious adverse events. In the arthroscopic surgery group, five of 164 (3%) participants had a serious knee-related adverse event (subsequent knee replacement) and three of 164 (1.8%) participants from this group had other serious adverse events (one acute myocardial infarction; one pulmonary embolism resulting in death; one vascular disorder), giving a total of eight of 164 (4.9%) participants with serious adverse events. At five years' follow-up, Katz 2013 reported that two of 109 (1.8%) participants randomised to and receiving physical therapy had a serious knee-related adverse event (subsequent knee replacement), and 16 of 164 participants randomised to and receiving arthroscopic surgery plus seven of 68 cross-over participants (for a total of 23 of 232 (10%) participants) had a serious knee-related adverse event (subsequent knee replacement). Other serious adverse events were not reported at five years so only the 12-month follow-up data could be included in Analysis 7.1.

Kise 2016 reported no serious knee-related or other serious adverse events in the exercise group. Kise 2016 reported that five of 64 (7.8%) participants randomised to and receiving arthroscopic surgery and two of 14 cross-over participants (for a total of seven of 78 (9%) participants) had serious knee-related adverse events (one knee replacement at 34 months; one osteotomy at four months; three partial meniscectomies of those randomised to arthroscopic surgery; one osteotomy at nine months; one partial meniscectomy). No other serious adverse events were reported with arthroscopic surgery.

Van de Graaf 2018 reported that eight of 162 (4.9%) participants had serious adverse events with physical therapy (one acute myocardial infarction; one sudden death; one neurological event; one alcoholic pancreatitis; three knee replacements; one arthroscopy). Van de Graaf 2018 reported that nine of 159 (5.7%) participants had serious adverse events with arthroscopic surgery (one intracranial malignancy; one lymph node malignancy; one rectal polyp; two knee replacements; two subsequent arthroscopies in affected knee; one arthroscopy in opposite knee; one other knee surgery). We included the apparent intervention-related adverse events in Analysis 7.1 (4/162 (2.5%) with physical therapy; 5/159 (3.1%) with arthroscopic surgery).

Merchan 1993 reported that two of 40 (5%) participants in the NSAIDs plus physiotherapy group had died after randomisation and that there were no other serious adverse events in this group. Merchan 1993 reported that seven of 40 (17.5%) participants in the arthroscopic surgery plus physiotherapy group had serious adverse events (five died after randomisation; two had deep vein thrombosis).

Vermesan 2013 reported that five study participants (4.2%) had serious knee-related adverse events (subsequent knee replacement) but the authors did not specify to which treatment group the participants belonged, so these data were not included in Analysis 7.1.

Total adverse events

Eleven trials reported total adverse events, defined as any adverse event, mild or serious in nature. This includes three placebo-controlled trials (Moseley 2002; Roos 2018; Sihvonen 2013), five trials with exercise control (Gauffin 2014; Herrlin 2007; Katz 2013; Kise 2016; Van de Graaf 2018), one trial with NSAIDs control (Merchan 1993), one trial with hyaluronic acid injection control (Saeed 2015), and one trial with a single intra-articular glucocorticoid injection control (Vermesan 2013).

Overall, total adverse events were reported in 95 of 634 (15%) participants in the control groups and in 114 of 692 (16.5%) participants in the arthroscopy groups from nine trials with usable data. Based on these trials, arthroscopic surgery may, or may not, slightly increase the risk of experiencing total adverse events at up

to five years compared with control (RR 1.15, 95% CI 0.78 to 1.70; I^2 = 48%; 9 trials, 1326 participants; low-certainty evidence; Analysis 7.2; Summary of findings 1). We downgraded the certainty of the evidence due to serious imprecision (small number of events) and possible reporting bias (incomplete reporting of outcomes across studies).

Roos 2018 reported that three of 22 (13.6%) participants in the placebo group and six of 22 (27.3%) in the arthroscopic surgery group had adverse events (including four knee-related and two other adverse events).

Sihvonen 2013 reported that eight of 76 (10.5%) participants from the placebo group had serious knee-related adverse events (one proximal tibial osteotomy; seven arthroscopic partial meniscectomies) and no other serious adverse events in this group. Sihvonen 2013 reported that seven of 70 (10%) participants from the arthroscopic surgery group had a serious knee-related adverse event (three knee replacements; 4 arthroscopies) and that one (1.4%) participant from this group had a serious adverse events (one deep infection at four months) (for a total of 8/70 (11.4%) participants).

Moseley 2002 reported that two participants had minor adverse events (incisional erythema treated with antibiotics; calf swelling but no thrombolysis on venography), but did not specify to which treatment group the participants belonged, so the data could not be included in Analysis 7.2.

Gauffin 2014 reported no adverse events in the exercise group, and that three of 66 (4.5%) participants in the arthroscopic surgery group had adverse events (two repeat arthroscopies; one death after three years).

Herrlin 2007 reported no serious adverse events with exercise. However, 13 of 49 (26.5%) participants crossed over from the exercise group and had an arthroscopic procedure on average 6.5 months following treatment, and three of 47 (6.4%) participants from the arthroscopy group underwent an additional arthroscopic procedure at between 13 and 40 months following their original surgery. No other adverse events were reported.

Katz 2013 reported that three of 109 (2.8%) participants from the physical therapy group had a knee-related adverse event (subsequent knee replacement) and that 15 of 109 (13.8%) participants from this group had other adverse events at 12 months (one stroke; one sudden death; one atrial fibrillation; one skin problem; four pain from fall or other trauma; one knee bursitis; one knee pain; four back/hip/foot pain; one other) (for a total of 18/109 (16.5%) participants). Katz 2013 reported that five of 164 (3%) participants from the arthroscopic surgery group had a knee-related adverse event (subsequent knee replacement) and that 18 of 164 (11%) participants from this group had other adverse events at 12 months (one acute myocardial infarction; one pulmonary embolism resulting in death; one hypoxaemia; two deep vein thrombosis; one syncope; two skin problems; two pain from fall or other trauma; three tendonitis; one rupture of Baker's cyst; one knee pain; two back/hip/foot pain; one other) (for a total of 23/164 (14%) participants). At five years' follow-up, Katz 2013 reported that two of 109 (1.8%) participants randomised to and receiving physical therapy had a knee-related adverse event (subsequent knee replacement) and that 16 of 164 participants randomised to and receiving arthroscopic surgery plus seven of 68 cross-over participants (for a total of 23/232 (9.9%) participants) had a knee-related adverse event (subsequent knee replacement). Other adverse events were not reported at five years so only the 12-month follow-up data could be included in Analysis 7.2.

Kise 2016 reported that 31 of 60 (51.7%) participants had adverse events with exercise (16 with pain, swelling, instability, stiffness or decreased range of motion in index knee and 15 with similar symptoms in the contralateral knee), and that 31 of 64 (48.4%) participants randomised to and receiving arthroscopic surgery and two of 14 cross-over participants (for a total of 33/78 (42.3%) participants) had adverse events (one knee replacement; one osteotomy; three partial meniscectomies; 16 knee pain, swelling, instability, stiffness or decreased range of motion; 10 with similar symptoms in the contralateral knee in those randomised to arthroscopic surgery; one osteotomy; one partial meniscectomy).

Van de Graaf 2018 reported that 12 of 162 (7.4%) participants had adverse events with physical therapy (one acute myocardial infarction, one sudden death, one neurological event; one alcoholic pancreatitis; three knee replacements; one arthroscopy; two knee pain resulting in extra consultations; two other musculoskeletal events). Van de Graaf 2018 reported that 18 of 159 (11.3%) participants had adverse events with arthroscopic surgery (one intracranial malignancy; one lymph node malignancy; one rectal polyp; two total knee replacements; two subsequent arthroscopies in affected knee; one arthroscopy in opposite knee; one other knee surgery; one reactive arthritis; six knee pain resulting in extra consultations; two back/hip/foot pain).

Merchan 1993 reported that two of 40 (5%) participants in the NSAIDs plus physiotherapy group had died after randomisation and no other adverse events. Merchan 1993 reported that nine of 40 (22.5%) participants in the arthroscopic surgery plus physiotherapy group had adverse events (five died after randomisation, two deep vein thrombosis, one superficial infection, one haemarthrosis).

Saeed 2015 reported that eight of 60 (13.3%) participants in the hyaluronic acid injection group had pain at the injection site, and that 13 of 60 (21.7%) participants in the arthroscopic surgery group experienced adverse events (pain and mild effusion) at six months' follow-up.

Vermesan 2013 reported that five study participants (4.2%) had a subsequent knee replacement but the authors did not specify to which treatment group the participants belonged, so the data could not be included in Analysis 7.2.

Adverse events in the surgery groups included seven deaths; one acute myocardial infarction; one hypoxaemia; four deep vein thrombosis; one intracranial malignancy; one lymph node malignancy; one syncope; one rectal polyp; one reactive arthritis; two skin problems; three tendonitis; two pain from fall or other trauma; one rupture of Baker's cyst; four back/hip/foot pain; and three other unspecified adverse events. Other knee-related adverse events included nine total knee replacements; two osteotomies; 13 repeat arthroscopies; one other knee surgery; one arthroscopy in opposite knee; one cutaneous nerve lesion; one deep infection; one superficial infection; 37 with knee pain, swelling, instability, stiffness or decreased range of motion; 10 with pain, swelling, instability, stiffness or decreased range of motion in the contralateral knee; and one haemarthrosis.

Adverse events in the control groups included four deaths; one acute myocardial infarction; one stroke; one neurological event (unspecified); one atrial fibrillation; one alcoholic pancreatitis; one skin problem; four pain from fall or other trauma; four back/hip/foot pain; and four other unspecified adverse events. Other knee-related adverse events included six knee replacements; one high tibial osteotomy; 14 arthroscopies; one knee bursitis; 27 with knee pain, swelling, instability, stiffness or decreased range of motion in index knee; 15 with pain, swelling, instability, stiffness or decreased range of motion in the contralateral knee; and two other unspecified musculoskeletal events.

Progression of knee osteoarthritis

Progression of knee osteoarthritis was reported in one placebo-controlled trial (Sihvonen 2013), and five exercise-controlled trials (Gauffin 2014; Herrlin 2007; Kise 2016; Van de Graaf 2018; Yim 2013).

In total, progression of knee osteoarthritis was reported in 69 of 256 (27.0%) participants in the control groups, and in 98 of 277 (35.4%) participants in the arthroscopic surgery groups from five trials where progression of knee osteoarthritis was reported as a dichotomous outcome and the data could be combined. Based on these studies, arthroscopic surgery may lead to greater progression of knee osteoarthritis at up to five years compared with control (RR 1.25, 95% CI 1.01 to 1.54; I^2 = 0%; 5 trials, 533 participants; low-certainty evidence; Analysis 7.3). We downgraded the certainty of the evidence two levels due to serious imprecision (small number of events).

Gauffin 2014 reported that 10 of 27 (37%) participants from the exercise group and 33 of 55 (60%) participants from the arthroscopy group had radiographic deterioration from baseline to the 5-year follow-up according to the Kellgren-Lawrence classification.

Herrlin 2007 reported that two of 45 (4.4%) participants from the exercise group and two of 43 (4.7%) participants from the arthroscopy group who underwent radiographic examination at the 5-year follow-up had evidence of osteoarthritis progression in the medial compartment from baseline, according to the Ahlback classification. In three cases, progression was from grade 1 to grade 2, and in one case from grade 1 to grade 3 (although the authors did not specify to which treatment group the participant with osteoarthritis progression from grade 1 to grade 3 belonged).

Kise 2016 reported that 10 of 58 (17.2%) participants from the exercise group and 13 of 62 (21.0%) participants from the arthroscopy group had radiographic knee osteoarthritis consistent with grade 2 or more on the Kellgren-Lawrence classification at five years.

Sihvonen 2013 reported that 44 of 74 (59.5%) participants from the placebo group and 48 of 67 (71.6%) participants from the arthroscopy group had at least one grade progression in radiographic tibiofemoral knee osteoarthritis on the Kellgren-Lawrence classification at five years.

Van de Graaf 2018 reported knee osteoarthritis severity progressed from 1.3 points at baseline to 1.5 points at 24 months on the Kellgren-Lawrence classification in the physical therapy group (MD 0.18 points, 95% CI 0.04 to 0.31), and from 1.3 points at baseline to 1.6 points at 24 months in the arthroscopy group (MD 0.37 points, 95% CI 0.25 to 0.49). The authors reported no significant between-group difference (0.10 points more progression in the arthroscopy group, 95% CI -0.05 to 0.26, P = 0.18).

Yim 2013 reported that three of 52 (5.8%) participants from the exercise group and two of 50 (4%) participants from the arthroscopy group had radiographic deterioration of grade 2 or more on the Kellgren-Lawrence classification at two years.

Subsequent knee surgery (replacement or high tibial osteotomy)

The need for subsequent knee surgery was reported in five trials: one placebo-controlled trial (Sihvonen 2013); three exercise-controlled trials (Katz 2013; Kise 2016; Van de Graaf 2018); and one glucocorticoid injection-controlled trial (Vermesan 2013).

In total, subsequent knee surgery (replacement or high tibial osteotomy) was reported in six of 407 (1.5%) participants in the control groups and in 23 of 457 (5%) participants in the arthroscopy groups from four trials with usable data. There is some imprecision due to the small event rate and the 95% confidence interval including both no difference between groups and a large increase in risk (RR 2.63, 95% CI 0.94 to 7.34; I^2 = 11%; 4 studies, 864 participants; low-certainty evidence; Analysis 7.4; Summary of findings 1). The certainty of the evidence was low, downgraded twice due to serious imprecision (small event rate).

Sihvonen 2013 reported that one of 76 (1%) participants from the placebo surgery group had a subsequent high tibial osteotomy and that three of 70 (4%) participants from the arthroscopic surgery group had a subsequent knee replacement at five years.

Katz 2013 reported that two of 109 (1.8%) randomised to and receiving physical therapy, and 16 of 164 (9.8%) randomised to and receiving arthroscopic surgery group plus seven of 68 (10.3%) cross-over participants from physical therapy to arthroscopy (for a total of 23/232 (10%) participants) had a subsequent total knee replacement.

Kise 2016 reported that zero of 60 participants randomised to and receiving exercise and two of 64 (3%) participants randomised to and receiving arthroscopic surgery plus one of 14 cross-over participants (for a total of 3 of 78 (4%) participants) had subsequent knee surgery (one total knee replacement at 34 months and one osteotomy at four months of those randomised to arthroscopic surgery, plus one osteotomy at nine months).

Van de Graaf 2018 reported that three of 162 (2%) participants from the exercise group and two of 159 (1%) participants from the arthroscopic surgery group had a subsequent knee replacement.

Vermesan 2013 reported that five study participants (4.2%) had a subsequent knee replacement, but the authors did not specify to which treatment group the participants belonged, so the data could not be included in Analysis 7.4.

DISCUSSION

Summary of main results

We found 16 randomised trials, including 2105 participants, that compared arthroscopic surgery with placebo surgery (4 trials, 380 participants) or non-surgical treatment (12 trials, 1725 participants) in people with degenerative knee disease, with or without degenerative meniscal tears. Our findings demonstrate that arthroscopic knee surgery provides little or no clinically

important benefit in pain or function in the short or longer term, and probably provides no clinically important benefit in knee-specific or generic quality of life, and may not improve treatment success, in the short or longer term compared with a placebo procedure (Summary of findings 1). Arthroscopic surgery may lead to little or no difference in, or slightly more, serious adverse events and total adverse events compared to placebo or non-surgical interventions. Few events may occur in this population even without surgery; yet most observed serious adverse events were likely attributable to the index procedure.

Arthroscopic surgery may lead to greater progression of knee osteoarthritis and it may or may not lead to a slight increase in subsequent knee surgery (replacement or osteotomy), although the 95% confidence intervals for subsequent knee surgery includes no difference between groups and a large increase in risk. We concluded that arthroscopy may increase the risk of osteoarthritis progression on the basis that the relative risk estimate indicated that there could be a 25% increased risk, although the 95% confidence interval included almost no increased risk to a greater than 50% increase in risk, and this result was consistent across studies (no statistical heterogeneity). For subsequent knee surgery, we concluded that there may or may not be an increased risk: while the relative risk estimate indicated that the increased risk could be 2.5 times greater and the upper bound of the 95% confidence interval indicated that there could be a large increase in risk, the lower bound of the 95% confidence interval also included no difference between groups (and there was likely no important heterogeneity across studies). However, the observed overall risk of progression of knee osteoarthritis and risk of subsequent knee surgery both increased by about 2% following arthroscopic surgery compared with control, estimates that are consistent with those reported in larger population-based studies (Winter 2017).

Compared to exercise, arthroscopic knee surgery probably provides little or no clinically important benefit in pain or function, and may provide little or no improvements in knee-specific and generic quality of life in the short or longer term. We are uncertain whether arthroscopic surgery leads to a difference in treatment success and progression of knee osteoarthritis because the certainty of the evidence is very low.

Compared to a single intra-articular glucocorticoid injection, arthroscopic knee surgery may slightly improve function in the short term but leads to little or no difference in the longer term. Pain, knee-specific and generic quality of life, treatment success and progression of knee osteoarthritis were not assessed in the one trial comparing these treatments.

Compared to non-arthroscopic lavage, arthroscopic knee surgery may lead to little or no difference in pain or function in the short or longer term, or treatment success. Knee-specific and generic quality of life, adverse events, progression of knee osteoarthritis and knee surgery (replacement or osteotomy) were not reported in the trial evaluating these treatments.

Compared to non-steroidal anti-inflammatory drugs (NSAIDs), arthroscopic knee surgery may improve participant-reported treatment success but its effect on function is uncertain because the certainty of the evidence is very low. Pain, knee-specific and generic quality of life, progression of knee osteoarthritis and knee surgery (replacement or osteotomy) were not assessed in the trial comparing these treatments.

It is uncertain whether arthroscopic knee surgery reduces pain in the short term compared to five hyaluronic acid injections given at weekly intervals because the certainty of the evidence is very low. Function, knee-specific quality of life, treatment success and progression of knee osteoarthritis were not assessed in the single trial comparing these treatments.

No trials compared arthroscopic surgery to complementary therapies.

Overall completeness and applicability of evidence

Overall, the trials included participants and interventions that are largely reflective of clinical practice, indicating that the results of the review can be broadly applied to practice. A quarter of studies were placebo-controlled trials with minimal biases, while half were open-label trials comparing arthroscopic surgery to common conservative care in the form of exercise or rehabilitation programmes. Two further trials compared surgery to other common conservative care modalities: intra-articular glucocorticoid injection and non-steroidal anti-inflammatory medication. Two trials used less common comparators: non-arthroscopic lavage and hyaluronic acid injections.

Trials were conducted across several countries, the majority in Europe, followed by North America, with single trials conducted in Pakistan and South Korea. The trials varied in their eligibility criteria regarding the presence of osteoarthritis and degenerative meniscal tears. For our primary comparison of arthroscopic surgery versus placebo surgery, half of the trials included participants with meniscal tears (and it was unclear if participants in the remaining trials had meniscal tears). Overall, trials included participants with symptoms of knee pain or torn menisci in the preceding few months. Some trials required evidence of degeneration or tears of the meniscus or cartilage on imaging or arthroscopy, and some trials included only participants with radiographic evidence of osteoarthritis, ranging from mild to severe disease. It is likely this variation reflects conditions seen in clinical practice.

Arthroscopic procedures varied from debridement of rough cartilage and trimming of torn/degenerated menisci to surgical resection of damaged menisci and excision of loose fragments of cartilage and bone. These procedures are reflective of those used in clinical practice.

The majority of trials were designed to assess the benefits of knee arthroscopy in terms of important patient-relevant outcomes: pain and knee function. Fewer trials (50%) measured treatment success and only five (31%) trials reported knee-specific quality of life, so we are less certain of the applicability of results for these outcomes.

We were unable to reliably estimate the harms associated with arthroscopic surgery from the included trials, as event rates were low, and trials were likely not large enough to detect important differences between groups. Our risk estimates were further hampered by the failure of nearly half the trials to report adverse events. Serious surgical-related adverse events may increase by an absolute risk of 2% with arthroscopic surgery, but due to the low number of events, there is some uncertainty around the estimate - the 95% confidence intervals include both an increase and small decrease in risk.

Longer-term outcomes, including progression of knee osteoarthritis and subsequent knee replacement, are also

important. Six trials reported progression of knee osteoarthritis and five measured knee surgery (replacement or osteotomy) as an outcome. These trials were not large or long enough to reliably assess these outcomes, but the absolute risk of subsequent knee replacement or knee surgery was about 2% greater following arthroscopic surgery than with control, which is consistent with the incidence of knee replacement after arthroscopic surgery reported in observational studies. A systematic review of 20 observational cohort and cross-sectional studies indicates the yearly incidence of knee replacement after arthroscopic surgery for osteoarthritis is 2.62% (95% CI 1.26% to 3.46%) and the median interval between arthroscopy and knee replacement is 2.0 years (Winter 2017). A long-term trial, an extension of Sihvonen 2013, is underway, with plans for a 10-year follow-up to assess the incidence of radiographically-confirmed osteoarthritis following knee arthroscopy, with results expected in 2024 (NCT01052233).

Quality of the evidence

Arthroscopic surgery versus placebo surgery

We did not downgrade the evidence for pain or function. We downgraded the evidence for knee-specific quality of life to moderate certainty due to serious imprecision, as the 95% confidence intervals did not rule in or rule out a clinically important change. We downgraded the evidence for participant-reported treatment success to low certainty due to serious indirectness - treatment success was defined variably across trials and measured at different time points - and serious imprecision, as there were small numbers reported.

We downgraded the evidence for serious adverse events and total adverse effects to low certainty due to serious imprecision (small number of events) and likely reporting bias (incomplete reporting of outcome across studies). Few events may occur in this population even without surgery; most observed serious adverse events were likely attributable to the index procedure.

We downgraded the evidence for both progression of knee osteoarthritis and subsequent knee surgery by two levels - to low certainty - due to the small number of events. The trials were not large or long enough to detect many events (98 with knee arthroscopy and 69 with control for progression of knee osteoarthritis; 23 with knee arthroscopy and 6 with control for subsequent knee surgery). Although only six and five studies reported progression of knee osteoarthritis and subsequent knee replacement, respectively, we did not downgrade for reporting bias as it appears the remaining studies did not intend to measure these outcomes.

Arthroscopic surgery versus exercise

We downgraded the evidence for pain and function to moderate certainty due to serious risk of bias (largely detection bias). We downgraded the evidence for knee-specific and generic quality of life to low certainty due to serious risk of bias and serious imprecision. We downgraded the evidence for treatment success to very low certainty due to serious risk of bias (all trials had high risk of performance and detection bias, and some concerns with selection bias), serious inconsistency and serious indirectness (diversity in outcome measurement across trials). We downgraded the evidence for progression of knee osteoarthritis to very low certainty due to very serious imprecision (very few events reported) and likely reporting bias.

Arthroscopic surgery versus glucocorticoid injection

We downgraded the evidence for function to low certainty due to serious risk of bias (some concerns with selection, performance and detection bias, high risk of reporting bias) and serious imprecision (only one study).

Arthroscopic surgery versus non-arthroscopic lavage

We downgraded the evidence for pain, function and treatment success to low certainty due to serious risk of bias (some concerns with selection and reporting bias, high risk of performance and detection bias) and serious imprecision (only one study).

Arthroscopic surgery versus anti-inflammatory drugs

We downgraded the evidence for treatment success to low certainty due to serious risk of bias (some concerns with selection and reporting bias, high risk of performance and detection bias) and serious imprecision (only one study with wide confidence intervals). We downgraded the evidence for function to very low certainty because of serious risk of bias, serious imprecision and reporting bias.

Arthroscopic surgery versus hyaluronic acid injections

We downgraded the evidence for pain to very low certainty due to serious risk of bias (some concerns with selection, performance and detection bias, high risk of reporting bias), serious imprecision (wide confidence intervals and only one study) and serious indirectness (outcome dissimilarity).

Potential biases in the review process

To the best of our knowledge, we identified all relevant trials meeting the review's eligibility criteria through a comprehensive search of major electronic databases and trial registries without language restrictions. We used up to three independent assessors to screen and select studies and perform risk of bias judgements. None of the review authors have been involved in the conduct of the included trials.

There were too few studies to formally assess the presence of publication bias. We identified four ongoing trials comparing arthroscopic surgery to exercise (NCT02113280; NCT02995551; NCT04837456; NCT04313569). As these trials are unblinded and thus subject to detection biases, it is unlikely the results, when available, will change the conclusions of this review for the comparison of arthroscopic surgery versus exercise.

Agreements and disagreements with other studies or reviews

In addition to our earlier systematic review (Brignardello-Petersen 2017), we identified 11 other systematic reviews (Abram 2020; Barlow 2015; Health Quality Ontario 2014; Hohmann 2018; Khan 2014; Lamplot 2016; Lee 2018; Li 2020; Monk 2017; Thorlund 2015; Van de Graaf 2018), and one narrative review (Ha 2016), assessing the effects of arthroscopic surgery compared to non-surgical interventions. Two focused exclusively on arthroscopic surgery versus exercise (Hohmann 2018; Li 2020), and eight included only studies with participants with meniscal tears (Abram 2020; Ha 2016; Hohmann 2018; Khan 2014; Lee 2018; Li 2020; Monk 2017; Van de Graaf 2016). Three reviews included studies other than randomised trials (Abram 2020; Lamplot 2016; Monk 2017); one excluded trials

because of no or insufficient data on pain or functional outcomes (Thorlund 2015); and eight reviews failed to identify one or more eligible trials that are included in this review update (Barlow 2015; Ha 2016; Health Quality Ontario 2014; Lamplot 2016; Lee 2018; Li 2020; Monk 2017; Van de Graaf 2016).

Abram 2020 included 10 trials that appear in this review update (Gauffin 2014; Herrlin 2007; Katz 2013; Kise 2016; Osteras 2012; Roos 2018; Sihvonen 2013; Van de Graaf 2018; Vermesan 2013; Yim 2013), and their conclusions are largely in agreement with this review update. They reported little or no difference in pain, function, knee-specific quality of life and generic quality of life with arthroscopic surgery compared to placebo in the short or longer term and no difference in presence of mechanical knee symptoms between groups. They reported low event rates for subsequent knee surgery (2.9% with arthroscopic surgery, 6.6% with placebo) and adverse events.

Lee 2018 included eight trials that appear in this review update (Chang 1993; Gauffin 2014; Herrlin 2007; Katz 2013; Kirkley 2008; Moseley 2002; Vermesan 2013; Yim 2013), but failed to identify three eligible trials at the time of their search (Kise 2016; Osteras 2012; Sihvonen 2013). They reported no significant differences between arthroscopic surgery and conservative management (including placebo surgery).

Monk 2017 included six trials that are included in this review update (Herrlin 2007; Katz 2013; Moseley 2002; Sihvonen 2013; Vermesan 2013; Yim 2013), but failed to identify Gauffin 2014 and Osteras 2012. They reported no clear benefit of arthroscopic surgery over non-operative treatment for participants with degenerative tears in the presence or absence of osteoarthritic changes.

Van de Graaf 2016 included five trials that are included in this review update (Herrlin 2007; Katz 2013; Osteras 2012; Sihvonen 2013; Yim 2013), but failed to identify Gauffin 2014 and Vermesan 2013. They reported a small statistically significant but clinically unimportant benefit in pain (MD 0.56, 95% CI 0.28 to 0.83 on VAS) and function (MD 3.56, 95% CI 0.24 to 6.88 on KOOS) with arthroscopic surgery compared with conservative treatment at up to six months, and no difference between groups at longer follow-up.

Lamplot 2016 included five trials included in this review update (Gauffin 2014; Katz 2013; Kirkley 2008; Merchan 1993; Sihvonen 2013), but failed to identify seven eligible trials at the time of their search that are included in this review update (Chang 1993; Herrlin 2007; Moseley 1996; Osteras 2012; Saeed 2015; Vermesan 2013; Yim 2013. The review authors vote counted based on statistical significance and reported two trials showed benefit with arthroscopic surgery compared to conservative treatment and three found no difference.

Ha 2016 included five trials that are included in this review update (Gauffin 2014; Herrlin 2007; Katz 2013; Sihvonen 2013; Yim 2013), but failed to identify Osteras 2012 and Vermesan 2013. The authors provided a narrative summary of individual trial findings (no synthesis) and said they could not draw any conclusions on the optimal treatment for meniscal tears.

Thorlund 2015 included eight trials that are included in this review update (Chang 1993; Herrlin 2007; Katz 2013; Kirkley 2008; Moseley 2002; Osteras 2012; Sihvonen 2013; Yim 2013). The authors reported a small clinically unimportant benefit in knee pain in the short term with arthroscopic surgery compared to control but no difference in the longer term or in function between groups. Harms reported included symptomatic deep venous thrombosis, pulmonary embolism, infection and death.

Barlow 2015 included three trials included in this review update (Chang 1993; Kirkley 2008; Moseley 2002), but failed to identify Merchan 1993 and Moseley 1996. The authors provided a narrative summary of individual trial findings and concluded that none of the trials support use of arthroscopy in people with osteoarthritis.

Khan 2014 included six trials included in this review update (Herrlin 2007; Katz 2013; Osteras 2012; Sihvonen 2013; Vermesan 2013; Yim 2013), and reported no clinically important difference in pain and function between arthroscopic surgery and non-operative or sham treatments in the short and longer term.

Health Quality Ontario 2014 included seven trials included in this review update (Herrlin 2007; Katz 2013; Kirkley 2008; Moseley 2002; Osteras 2012; Sihvonen 2013; Yim 2013), but failed to identify four eligible trials at the time of their search (Chang 1993; Merchan 1993; Moseley 1996; Vermesan 2013). They reported no difference in pain or function between arthroscopic surgery and placebo, and between arthroscopic surgery and usual care (e.g. physical therapy).

Li 2020 included six trials included in this review update (Herrlin 2007; Katz 2013; Kise 2016; Osteras 2012; Van de Graaf 2018; Yim 2013), but failed to identify Gauffin 2014. They reported a small benefit in pain and function with arthroscopic surgery compared with physical exercise in the short term but no difference between groups in the longer term. Li 2020 reported no difference in osteoarthritis progression between arthroscopic surgery and exercise based on two trials (Herrlin 2007; Van de Graaf 2018).

Hohmann 2018 included six trials included in this review update (Gauffin 2014; Herrlin 2007; Katz 2013; Kise 2016; Osteras 2012; Yim 2013), and concluded that there is no compelling evidence to support arthroscopic surgery compared to physical therapy.

Only Li 2020 assessed the effect of arthroscopic surgery on the progression of knee osteoarthritis compared with exercise, but failed to report on this outcome in the results of the review.

AUTHORS' CONCLUSIONS

Implications for practice

The findings of this review demonstrate that arthroscopic surgery for people with symptomatic degenerative knee disease (average age ranging from 46 to 65 years; 56% women), provides little or no clinically important benefit in pain or function, probably does not provide clinically important benefits in knee-specific quality of life, and may not improve treatment success compared with placebo surgery. These results apply to people with knee osteoarthritis with or without meniscal tears, as well as people with meniscal tears alone. Arthroscopic surgery may or may not increase serious and total adverse events compared to control. Serious adverse events include deep venous thrombosis, myocardial infarction, pulmonary embolism, infection and death, and are likely mostly attributable to the arthroscopic surgery. Arthroscopic surgery may lead to greater progression of knee osteoarthritis and may or may not slightly increase subsequent knee surgery (replacement or osteotomy).

Participants in the included trials experienced improvement in pain and function over time whether or not they received surgery, placebo surgery or other control treatment. People contemplating arthroscopic surgery should be informed about the findings of this evidence synthesis to help them make an evidence-informed decision. They should also be informed that their symptoms are likely to improve slowly over time irrespective of treatment and that surgery has the potential for short-term harms related to the surgery and long-term harms of greater progression of knee osteoarthritis and the need for further knee surgery.

Implications for research

Given there are no benefits in pain and function, probably no benefit in quality of life, and maybe no difference in treatment success with arthroscopic surgery compared to placebo surgery, more randomised placebo-controlled trials assessing benefits of knee arthroscopic surgery are likely unnecessary. If proponents of the procedure consider there may still be one or more subgroups who may benefit from arthroscopic surgery, then the onus is on them to provide this evidence. However, to date, several studies have failed to find evidence of these subgroups (Sihvonen 2018; Pihl 2020).

We are less certain if arthroscopic surgery leads to more serious and total adverse events, earlier progression of knee osteoarthritis and more knee surgery (replacement or osteotomy) compared with placebo. Longer-term follow-up of participants in the included placebo-controlled trials or data from prospectively designed registries would provide more precise estimates of the risk of adverse events, progression of knee osteoarthritis and subsequent knee replacement with arthroscopic surgery. However, given there is high-certainty evidence of no benefit, the value of assessing if there is more harm may be limited.

Future updates of this review may be considered if further placebo-controlled trials are published that are likely to increase the certainty of effect estimates of harms.

Further trials comparing arthroscopic surgery with non-surgical interventions are likely to be of limited value, and we are unlikely to include such studies in future updates.

ACKNOWLEDGEMENTS

We thank the authors of the following articles, who supplied data missing from their published articles, upon request: Herrlin 2007; Kise 2016; Roos 2018; Sihvonen 2013.

This is an update of a previous review (Brignardello-Petersen 2017), and we acknowledge the authors of that review: Romina Brignardello-Petersen, Gordon Guyatt, Rachelle Buchbinder, Rudolf Poolman, Stefan Schandelmaier, Yaping Chang, Behnam Sadeghirad, Nathan Evaniew and Per Vandvik.

We acknowledge peer reviewers Dr Aleksi Reito, Department of Musculoskeletal Diseases, Tampere University Hospital, School of Medicine and Health Technology, Tampere University, Finland; Professor of Orthopaedic Surgery Ian Harris, University of New South Wales, Sydney; consumer reviewer Ms Catherine Hofstetter, and copy editor Faith Armitage.

REFERENCES

References to studies included in this review

Chang 1993 *{published and unpublished data}*

Chang RW, Facloner J, Stulberg SD, Arnold WJ, Dyer AR.Prerandomization: an alternative to classic randomization. *Journal of Bone and Joint Surgery* 1990;**72-A**:1451-5.

* Chang RW, Falconer J, Stulberg SD, Arnold WJ, Manheim LM, Dyer AR.A randomized, controlled trial of arthroscopic surgery versus closed-needle joint lavage for patients with osteoarthritis of the knee. *Arthritis and Rheumatism* 1993;**36**(3):289–96.

Gauffin 2014 *{published and unpublished data}*

Gauffin H, Kvist J, Hedevik H, Sonesson S.Knee arthroscopic surgery in middle-aged patients with meniscal symptoms: a 5-year follow-up of a prospective, randomized study. *Osteoarthritis and Cartilage* 2019;**27**:S-231.

Gauffin H, Sonesson S, Meunier A, Magnusson H, Kvist J.Knee arthroscopic surgery in middle-aged patients with meniscal symptoms: a 3-year follow-up of a prospective, randomized study. *American Journal of Sports Medicine* 2017;**45**(9):2077-84.

* Gauffin H, Tagesson S, Meunier A, Magnusson H, Kvist J.Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: a prospective, randomised, single-blinded study. *Osteoarthritis and Cartilage* 2014;**22**:1808–16.

Sonesson S, Kvist J, Yacob J, Hedevik H, Gauffin H.Knee arthroscopic surgery in middle-aged patients with meniscal symptoms: a 5-year follow-up of a prospective, randomized study. *Orthopaedic Journal of Sports Medicine* 2020;**8**(1):[12 p.].

Herrlin 2007 *{published and unpublished data}*

* Herrlin S, Hallander M, Wange P, Weidenhielm L, Werner S.Arthroscopic or conservative treatment of degenerative medial meniscal tears: a prospective randomised trial. *Knee Surgery, Sports Traumatology, Arthroscopy* 2007;**15**:393–401.

Herrlin SV, Wange PO, Lapidus G, Hallander M, Werner S, Weidenhielm L.Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. *Knee Surgery, Sports Traumatology, Arthroscopy* 2013;**21**(2):358-364.

Katz 2013 *{published and unpublished data}*

Collins JE, Losine E, Marx RG, Guermazi A, Jarraya M, Jones MH, et al, MeTeOR Investigator Group.Early magnetic resonance imaging-based changes in patients with meniscal tear and osteoarthritis: eighteen-month data from a randomized controlled trial of arthroscopic partial meniscectomy versus physical therapy. *Arthritis Care & Research* 2020;**72**:630-40.

* Katz JN, Brophy RH, Chaisson CE, de Chaves L, Cole BJ, Dahm DL, et al.Surgery versus physical therapy for a meniscal tear and osteoarthritis. *New England Journal of Medicine* 2013;**368**(18):1675–84. [NCT00597012]

Katz JN, Chaisson CE, Cole B, Guermazi A, Hunter DJ, Jones M, et al.The MeTeOR trial (Meniscal Tear in Osteoarthritis Research): rationale and design features. *Contemporary Clinical Trials* 2012;**33**(6):1189-96. [NCT00597012]

Katz JN, Shrestha S, Losina E, Jones MH, Marx RG, Mandl LA, et al, METEOR Investigators.Five-year outcome of operative and nonoperative management of meniscal tear in persons older than forty-five years. *Arthritis and Rheumatology* 2020;**72**(2):273-81.

Katz JN, Shrestha S, Losina E, Jones MH, Marx RG, Mandl LA, METEOR Investigators.Five-year outcome of operative and non-operative management of meniscal tear in persons greater than 45 years old. *Arthritis and Rheumatology* 2020;**72**(2):273-81.

Katz JN, Shrestha S, Losina E, Mandl LA, Levy B, Spindler K, et al.Five-year outcome of operative and nonoperative management of meniscal tear in the presence of osteoarthritic changes [abstract]. *Arthritis and Rheumatology* 2018;**70**(Suppl 10):Abstract no. 1816.

Katz JN, Spindler K, Safran-Norton C, Martin S, Mandl L, Jones M, et al.Predictors and outcomes of cross-over to surgery in a randomized trial of surgery vs. physical therapy for meniscal tear and osteoarthritis. *Osteoarthritis and Cartilage* 2015;**23**(Suppl 2):A36.

Shrestha S, Katz J, Losina E, Collins J.Five year structural changes in patients with meniscal tear and osteoarthritis from an RCT of arthroscopic partial meniscectomy vs. physical therapy. *Arthritis and Rheumatology* 2019;**71**(Suppl 10):4975-77.

Kirkley 2008 *{published and unpublished data}*

* Kirkley A, Birmingham TB, Litchfield RB, Giffin JR, Willits KR, Wong CJ, et al.A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine* 2008;**359**(11):1097–107. Erratum in: New England Journal of Medicine 2009; 361(20):2004.

Marsh J, Birmingham TB, Giffin JR, Isaranuwatchai W, Hoch JS, Litchfield R, et al.Cost-effectiveness analysis of arthroscopic surgery compared to non-operative management for osteoarthritis of the knee. *Osteoarthritis and Cartilage* 2015;**23**(Suppl 2):A31.

Marsh JD, Birmingham TB, Giffin JR, Isaranuwatchai W, Hoch JS, Feagan BG, et al.Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. *BMJ Open* 2016;**6**(1):e009949. [NCT00158431]

Kise 2016 *{published and unpublished data}*

Berg B, Roos EM, Englund M, Kise NJ, Tiulpin A, Saarakkala S, et al.Development of osteoarthritis in patients with degenerative meniscal tears treated with exercise therapy or surgery: a randomized controlled trial. *Osteoarthritis and Cartilage* 2020;**28**(7):897-906.

Berg B, Roos EM, Englund M, Kise NJ, Tiulpin A, Saarakkala S, et al.Knee osteoarthritis development five years following

arthroscopic partial meniscectomy or exercise therapy for degenerative meniscal tears: the Odense-Oslo meniscectomy versus exercise trial. *Osteoarthritis and Cartilage* 2020;**28**(Suppl):S28. [DOI: 10.1016/j.joca.2020.02.045]

* Kise NJ, Risberg MA, Stensrud S, Ranstam J, Engebretsen L, Roos EM.Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. *BMJ* 2016;**354**:i3740. [NCT01002794]

Stensrud S, Risberg MA, Roos EM.Effect of exercise therapy compared with arthroscopic surgery on knee muscle strength and functional performance in middle-aged patients with degenerative meniscus tears: a 3-mo follow-up of a randomized controlled trial. *American Journal of Physical Medicine and Rehabilitation* 2015;**94**(6):460–73. [NCT01002794]

Merchan 1993 *{published and unpublished data}*

* Merchan EC and Galindo E.Arthroscope-guided surgery versus nonoperative treatment for limited degenerative osteoarthritis of the femorotibial joint in patients over 50 years of age: a prospective comparative study. *Arthroscopy* 1993;**9**(6):663-7.

Moseley 1996 *{published data only}*

* Moseley JB Jr, Wray NP, Kuykendall D, Willis K, Landon G.Arthroscopic treatment of osteoarthritis of the knee: a prospective, randomized, placebo-controlled trial. Results of a pilot study. *American Journal of Sports Medicine* 1996;**24**(1):28-34.

Moseley 2002 *{published data only}*

* Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, et al.A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine* 2002;**347**(2):81-8.

Osteras 2012 *{published and unpublished data}*

* Osteras H, Osteras B, Torstensen TA.Medical exercise therapy, and not arthroscopic surgery, resulted in decreased depression and anxiety in patients with degenerative meniscus injury. *Journal of Bodywork and Movement Therapies* 2012;**16**(4):456–63.

Osteras H, Torstensen TA, Selven E, Haugerud L.High dosage medical exercise therapy or arthroscopic treatment for patients with degenerative meniscus injury: a pilot study. *Physiotherapy* 2011;**97**(Suppl 1):eS946-7.

Roos 2018 *{published data only}*

Hare KB, Lohmander LS, Christensen R, Roos EM.Arthroscopic partial meniscectomy in middle-aged patients with mild or no knee osteoarthritis: a protocol for a double-blind, randomized sham-controlled multi-centre trial. *BMC Musculoskeletal Disorders* 2013;**14**:71. [NCT01264991]

Hare KB, Lohmander LS, Roos EM.The challenge of recruiting patients into a placebo-controlled surgical trial. *Trials* 2014;**15**:167. [NCT01264991]

Hare KB, Lohmander S, Roos EM.The challenges of recruiting patients into a sham surgery trial. *Osteoarthritis and Cartilage* 2012;**20**:S173-4. [NCT01264991]

* Roos EM, Hare KB, Nielsen SB, Christensen R, Lohmander LS.Better outcome from arthroscopic partial meniscectomy than skin incisions only? A sham-controlled randomised trial in patients aged 35–55 years with knee pain and an MRI-verified meniscal tear. *BMJ Open* 2018;**8**(2):[9 p.]. [DOI: doi:10.1136/bmjopen-2017-019461] [NCT01264991]

Saeed 2015 *{published data only}*

Saeed K, Khan SA, Ahmed I.Efficacy of intra articular hyaluronic acid versus arthroscopic debridement in terms of improvement in pain score in Kellgran-Lawrence Grading II & III osteoarthritis of knee joint. *Pakistan Journal of Medical and Health Sciences* 2015;**9**(3):1011–5.

Sihvonen 2013 *{published and unpublished data}*

Jarvinen T, Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, et al.Arthroscopic partial meniscectomy vs sham surgery for degenerative meniscus tear. *Arthroscopy* 2014;**30**(6 Supp 1):e38.

Sihvonen R, Englund M, Turkiewicz A, Järvinen TL, Finnish Degenerative Meniscal Lesion Study Group.Mechanical symptoms and arthroscopic partial meniscectomy in patients with degenerative meniscus tear: a secondary analysis of a randomized trial. *Annals of Internal Medicine* 2016;**164**(7):449-55. [DOI: 10.7326/M15-0899] [NCT00549172]

Sihvonen R, Kalske R, Englund M, Turkiewics A, Toivonen P, Taimela S, Finnish Degenerative Meniscal Lesion Study (FIDELITY) Investigators.Statistical analysis plan for the 5-year and 10-year follow-up assessments of the FIDELITY trial: study protocol. *Trials* 2020;**21**:76. [NCT01052233]

Sihvonen R, Paavola M, Malmivaara A, Itälä A, Joukainen A, Kalske J, FIDELITY (Finnish Degenerative Meniscus Lesion Study) Investigators.Arthroscopic partial meniscectomy for a degenerative meniscus tear: a 5 year follow-up of the placebo-surgery controlled FIDELITY (Finnish Degenerative Meniscus Lesion Study) trial. *British Journal of Sports Medicine* 2020;**54**(22):1332-9. [DOI: 10.1136/bjsports-2020-102813<]

Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H, et al.Arthroscopic partial meniscectomy versus placebo surgery for a degenerative meniscus tear: a 2-year follow-up of the randomised controlled trial. *Annals of the Rheumatic Diseases* 2018;**77**:188-95.

* Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H, et al.Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *New England Journal of Medicine* 2013;**369**(26):2515–24. [NCT00549172]

Sihvonen R, Paavola M, Malmivaara A, Järvinen TL.Finnish Degenerative Meniscal Lesion Study (FIDELITY): a protocol for a randomised, placebo surgery controlled trial on the efficacy of arthroscopic partial meniscectomy for patients with degenerative meniscus injury with a novel 'RCT within-a-cohort' study design. *BMJ Open* 2013;**3**(3):e002510. [DOI: 10.1136/bmjopen-2012-002510] [NCT00549172]

Van de Graaf 2018 *{published data only}*

Noorduyn JC, Glastra van Loon T, Van de Graaf VA, Willigenburg NW, Butter IK, Scholten-Peeters GG, et al.Functional outcomes of arthroscopic partial meniscectomy versus physical therapy for degenerative meniscal tears using a patient-specific score: a randomized controlled trial. *Orthopaedic Journal of Sports Medicine* 2020;**8**(10):1-12.

Van de Graaf VA, Noorduyn JC, Willigenburg NW, et al.Effect of early surgery vs physical therapy on knee function among patients with nonobstructive meniscal tears: the ESCAPE randomized clinical trial. *JAMA* 2018;**320**(13):1328–37. [DOI: 10.1001/jama.2018.13308]

Van de Graaf VA, Scholtes VA, Wolterbeek N, Noorduyn JC, Neeter C, Van Tulder MW, et al, Escape Research Group.Cost-effectiveness of early surgery versus conservative treatment with optional delayed meniscectomy for patients over 45 years with non-obstructive meniscal tears (ESCAPE study): protocol of a randomised controlled trial. *BMJ Open* 2016;**6**:e014381.

Van de Graaf VA, Van Dongen JM, Willigenburg NW, Noorduyn JC, Butter IK, De Gast A, et al, ESCAPE Research Group.How do the costs of physical therapy and arthroscopic partial meniscectomy compare? A trial-based economic evaluation of two treatments in patients with meniscal tears alongside the ESCAPE study. *British Journal of Sports Medicine* 2020;**54**:538-46.

Vermesan 2013 *{published and unpublished data}*

* Vermesan D, Prejbeanu R, Laitin S, Damian G, Deleanu B, Abbinante A, et al.Arthroscopic debridement compared to intra-articular steroids in treating degenerative medial meniscal tears. *European Review for Medical and Pharmacological Sciences* 2013;**17**:3192–6.

Yim 2013 *{published and unpublished data}*

* Yim JH, Seon JK, Song EK, Choi JL, Kim MC, Lee KB, et al.A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. *American Journal of Sports Medicine* 2013;**41**(7):1565–70.

References to studies excluded from this review

Ahn 2015 *{published data only}*

Ahn JH, Jeong HJ, Lee YS, Park JH, Lee JW, Park J-H, et al.Comparison between conservative treatment and arthroscopic pull-out repair of the medial meniscus root tear and analysis of prognostic factors for the determination of repair indication. *Archives of Orthopaedic and Trauma Surgery* 2015;**135**:1265–76.

Biedert 2000 *{published data only}*

Biedert RM.Treatment of intrasubstance meniscal lesions: a randomized prospective study of four different methods. *Knee Surgery, Sports Traumatology, Arthroscopy* 2000;**8**(2):104-8. [PMID: 10795673]

Bisson 2015 *{published data only}*

* Bisson LJ, Kluczynski MA, Wind WM, Fineberg MS, Bernas GA, Rauh MA, et al.Design of a randomized controlled trial to compare debridement to observation of chondral lesions encountered during partial meniscectomy: the ChAMP (Chondral Lesions And Meniscus Procedures) trial. *Contemporary Clinical Trials* 2015;**45**(Pt B):281-6.

Bisson LJ, Kluczynski MA, Wind WM, Fineberg MS, Bernas GA, Rauh MA, et al.Patient outcomes after observation versus debridement of unstable chondral lesions during partial meniscectomy. *Journal of Bone and Joint Surgery* 2017;**99**(13):1078-85.

Bradley 2002 *{published data only}*

Bradley JD, Heilman DK, Katz BP, Gsell P, Wallick JE, Brandt KD.Tidal irrigation as treatment for knee osteoarthritis: a sham-controlled, randomized, double-blinded evaluation. *Arthritis and Rheumatology* 2002;**46**(1):100-8.

Hubbard 1996 *{published data only}*

Hubbard MJ.Articular debridement versus washout for degeneration of the medial femoral condyle. *Journal of Bone and Joint Surgery* 1996;**78**:B:217-9.

Kalunian 2000 *{published data only}*

Kalunian KC, Moreland LW, Klashman DJ, Brion PH, Concoff AL, Myers S, et al.Visually-guided irrigation in patients with early knee osteoarthritis: a multicenter randomized, controlled trial. *Osteoarthritis and Cartilage* 2000;**8**:412–8.

Lee 2020 *{published data only}*

Lee SH, Lee OS, Kim ST, Lee YS.Revisiting arthroscopic partial meniscectomy for degenerative tears in knees with mild or no osteoarthritis: a systematic review and meta-analysis of randomized controlled trials. *Clinical Journal of Sports Medicine* 2020;**30**(3):195-202. [DOI: 10.1097/JSM.0000000000000585]

Lu 2018 *{published data only}*

Lu Y-C, Bi B, Xiang Y-S, Du X-Y.Effectiveness and safety of arthroscopic debridement for treatment of degenerative knee osteoarthritis in elderly patients: study protocol for a non-randomized controlled trial. *Clinical Trials in Degenerative Diseases* 2018;**3**(1):15-21.

Ma 2020 *{published data only}*

Ma J, Chen H, Liu A, Cui Y, Ma X.Medical exercise therapy alone versus arthroscopic partial meniscectomy followed by medical exercise therapy for degenerative meniscal tear: a systematic review and meta-analysis of randomized controlled trials. *Journal of Orthopaedic Surgery and Research* 2020;**15**(1):219. [DOI: 10.1186/s13018-020-01741-3]

Marsh 2016 *{published data only}*

Marsh JD, Birmingham TB, Giffin JR, Isaranuwatchai W, Hoch JS, Feagan BG, et al.Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. *BMJ Open* 2016;**6**(5):e009949. [DOI: 10.1136/bmjopen-2015-009949]

Pan 2020 *{published data only}*

Pan H, Zhang P, Zhang Z, Yang Q.Arthroscopic partial meniscectomy combined with medical exercise therapy versus isolated medical exercise therapy for degenerative meniscal tear: a meta-analysis of randomized controlled trials. *International Journal of Surgery* 2020;**79**:222-32. [DOI: 10.1016/j.ijsu.2020.05.035]

Rimington 2009 *{published data only}*

Rimington T, Mallik K, Evans D, Mroczek K, Reider B.A prospective study of the non-operative treatment of degenerative meniscus tears. *Orthopedics* 2009;**32**(8):[no pagination]. [DOI: 10.3928/01477447-20090624-06]

Wijn 2020 *{published data only}*

Wijn SR, Rovers MM, Rongen JJ, Østerås H, Risberg MA, Roos EM, et al.Arthroscopic meniscectomy versus non-surgical or sham treatment in patients with MRI confirmed degenerative meniscus lesions: a protocol for an individual participant data meta-analysis. *BMJ Open* 2020;**10**:e031864. [DOI: 10.1136/bmjopen-2019-031864]

Zhang 2018 *{published data only}*

Zhang Y-F, Liu H.Clinical efficacy of knee arthroscopy in the treatment of degenerative knee osteoarthritis. *Biomedical Research* 2018;**29**(5):958-61. [DOI: 10.4066/biomedicalresearch.29-17-3351]

Zhao 2018 *{published data only}*

Zhao B, Yu Y, Liu W, Du J.Efficacy of arthroscopic loose body removal for knee osteoarthritis. *Experimental and Therapeutic Medicine* 2018;**15**(2):1666-71.

References to studies awaiting assessment

Kang 2005 *{published data only}*

Kang JG.Treatment of knee osteoarthritis with arthroscopic debridement and intra-articular sodium hyaluronate injection. *Journal of Jilin University Medicine Edition* 2005;**31**(5):802-5.

NCT00562822 *{published data only}*

NCT00562822.Surgery versus no surgery for osteoarthritis (OA) of the knee (MRC Knee). clinicaltrials.gov/ct2/show/NCT00562822 (first received 27 November 2007). [NCT00562822]

References to ongoing studies

NCT02113280 *{published data only}*

NCT02113280.DEMAND - DEgenerative Meniscal Tears - Arthroscopy vs. Dedicated Exercise (DEMAND). clinicaltrials.gov/ct2/show/NCT02113280 (first received 14 April 2014). [NCT02113280]

NCT02995551 *{published data only}*

NCT02995551.Danish RCT on Exercise versus Arthroscopic Meniscal surgery for young adults (DREAM). clinicaltrials.gov/ct2/show/NCT02995551 (first received 16 December 2016). [NCT02995551]

Skou ST, Lind M, Hölmich P, Jensen HP, Jensen C, Afzal M, et al.Study protocol for a randomised controlled trial of meniscal surgery compared with exercise and patient education for treatment of meniscal tears in young adults. *BMJ Open* 2017;**7**(8):e017436. [DOI: 10.1136/bmjopen-2017-017436]

NCT04313569 *{published data only}*

NCT04313569.Arthroscopic versus conservative treatment of degenerative meniscal tear in middle aged patients in regard to pain and knee function. clinicaltrials.gov/ct2/show/NCT04313569 (first received 18 March 2020).

NCT04837456 *{published data only}*

NCT04837456.Metabolic syndrome and degenerate meniscus tears. clinicaltrials.gov/ct2/show/NCT04837456 (first received 8 April 2021).

Additional references

Abram 2019a

Abram SG, Beard DJ, Price AJ, on behalf of the BASK Meniscal Working Group.Arthroscopic meniscal surgery: a national society treatment guideline and consensus statement. *Bone and Joint Journal* 2019;**101-B**:652-9.

Abram 2019b

Abram SG, Judge A, Beard D, Price A.Rates of knee arthroplasty within one-year of undergoing arthroscopic partial meniscectomy in England: temporal trends, regional and age-group variation in conversion rate. *Osteoarthritis and Cartilage* 2019;**27**(10):1420-9.

Abram 2020

Abram SG, Hopewell S, Monk AP, Bayliss LE, Beard DJ, Price AJ.Arthroscopic partial meniscectomy for meniscal tears of the knee: a systematic review and meta-analysis. *British Journal of Sports Medicine* 2020;**54**(11):652-63.

ACQSHC 2017

Australian Commission on Safety and Quality in Health Care.Osteoarthritis of the Knee Clinical Care Standard. Sydney, Australia: ACSQHC 2017.

Ahlback 1968

Ahlback S.Osteoarthritis of the knee. A radiographic investigation. *Acta Radiologica* 1968;**277**:7–72.

Alkan 2014

Alkan BM, Fidan F, Tosun A, Ardıçoğlu O.Quality of life and self-reported disability in patients with knee osteoarthritis. *Modern Rheumatology* 2014;**24**:166–71.

Australian Knee Society 2016

Australian Knee Society.Position Statement from the Australian Knee Society on Arthroscopic Surgery of the Knee, including reference to the presence of Osteoarthritis or Degenerative Joint Disease. Australian Orthopaedic Association October 2016.

Barlow 2015

Barlow T, Downham C, Griffin D.Arthroscopy in knee osteoarthritis: a systematic review of the literature. *Acta Orthopaedica Belgica* 2015;**81**(1):1-8.

Bartel 2016

Bartels EM, Juhl CB, Christensen R, Hagen KB, Danneskiold-Samsøe B, Dagfinrud H, et al.Aquatic exercise for the treatment of knee and hip osteoarthritis. *Cochrane Database of Systematic Reviews* 2016, Issue 3. Art. No: CD005523. [DOI: 10.1002/14651858.CD005523.pub3]

Bhattacharyya 2003

Bhattacharyya T, Gale D, Dewire P, Totterman S, Gale ME, McLaughlin S, et al.The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. *Journal of Bone and Joint Surgery. American Volume* 2003;**85**:4-9.

Brignardello-Petersen 2017

Brignardello-Petersen R, Guyatt GH, Buchbinder R, Poolman RW, Schandelmaier S, Chang Y, et al.Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. *BMJ Open* 2017;**7**:e016114. [DOI: 10.1136/bmjopen-2017-016114]

Brouwer 2014

Brouwer RW, Huizinga MR, Duivenvoorden T, Van Raaij TM, Verhagen AP, Bierma-Zeinstra SM, et al.Osteotomy for treating knee osteoarthritis.. *Cochrane Database of Systematic Reviews* 2014, Issue 12. Art. No: CD004019. [DOI: 10.1002/14651858.CD004019.pub4]

Brown 2013

Brown GA.AAOS clinical practice guideline: treatment of osteoarthritis of the knee: evidence-based guideline, 2nd edition. *Journal of the American Academy of Orthopedic Surgeons* 2013;**21**:577–9.

Buchbinder 2015

Buchbinder R, Harris IA, Sprowson A.Management of degenerative meniscal tears and the role of surgery. *BMJ* 2015;**350**:h2212. [DOI: 10.1136/bmj.h2212]

Cates 2008 [Computer program]

Visual Rx.Cates C, Version 3. Dr. Christopher Cates EBM website, 2008. Available at www.nntonline.net.

Christensen 2007

Christensen R, Bartels EM, Astrup A, Bliddal H.Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. *Annals of the Rheumatic Diseases* 2007;**66**(4):433-439.

Dawson 1998

Dawson J, Fitzpatrick R, Murray D, Carr A.Questionnaire on the perceptions of patients about total knee replacement. *Journal of Bone and Joint Surgery. British Volume* 1998;**80**(1):63-9.

Dearing 2010

Dearing J, Brenkel I.Incidence of knee arthroscopy in patients over 60 years of age in Scotland. *Surgeon* 2010;**8**:144-50.

Deeks 2021

Deeks JJ, Higgins JP, Altman DG (editors).Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston MS, Li T, Page MJ, et al (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 6.2 (updated February 2021), Cochrane, 2021. Available from www.training.cochrane.org/handbook/archive/v6.2/chapter-10.

Devji 2017

Devji T, Guyatt GH, Lytvyn L, Brignardello-Petersen R, Foroutan F, Sadeghirad B, et al.Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. *BMJ Open* 2017;**7**:e015587.

Englund 2008

Englund M, Guermazi A, Gale D, Hunter DJ, Aliabadi P, Clancy M, et al.Incidental meniscal findings on knee MRI in middle-aged and elderly persons. *New England Journal of Medicine* 2008;**359**:1108-15.

Fransen 2015

Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL.Exercise for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews* 2015, Issue 1. Art. No: CD004376. [DOI: 10.1002/14651858.CD004376.pub3]

Ghogomu 2014

Ghogomu EA, Maxwell LJ, Buchbinder R, Rader T, Pardo Pardo J, Johnston RV, et al.Updated method guidelines for Cochrane musculoskeletal group systematic reviews and meta-analyses. *Journal of Rheumatology* 2014;**41**(2):194-205.

GRADEPro GDT [Computer program]

McMaster University (developed by Evidence Prime) GRADEpro GDT.Version accessed prior to 14 February 2022. Hamilton (ON): McMaster University (developed by Evidence Prime), 2015. Available from www.gradepro.org.

Guermazi 2012

Guermazi A, Niu J, Hayashi D, Roemer FW, Englund M, Neogi T, et al.Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). *BMJ* 2012;**345**:e5339.

Ha 2016

Ha AY, Shalvoy RM, Voisinet A, Racine J, Aaron RK.Controversial role of arthroscopic meniscectomy of the knee: a review. *World Journal of Orthopedics* 2016;**7**(5):287-92.

Harris 2013

Harris IA, Madan NS, Naylor JM, Chong S, Mittal R, Jalaludin BB.Trends in knee arthroscopy and subsequent arthroplasty in an Australian population: a retrospective cohort study. *BMC Musculoskeletal Disorders* 2013;**14**:143.

Hawker 2008

Hawker G, Guan J, Judge A, Dieppe P.Knee arthroscopy in England and Ontario: patterns of use, changes over time, and relationship to total knee replacement. *Journal of Bone and Joint Surgery* 2008;**90**:2337-45.

Health Quality Ontario 2014

Evidence Development and Standards Branch, Health Quality Ontario.Arthroscopic debridement of the knee: an evidence update. Ontario Health Technology Assessment Series 2014;**14**(13):1-43.

Hegedus 2007

Hegedus EJ, Cook C, Hasselblad V, Goode A, McCrory DC.Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. *Journal of Orthopaedic and Sports Physical Therapy* 2007;**37**:541-50.

Higgins 2017

Higgins JP, Altman DG, Sterne JA, editor(s).Chapter 8: Assessing risk of bias in included studies. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor(s), Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated June 2017). Cochrane, 2017. Available from www.training.cochrane.org/handbook.

Higgins 2021

Higgins JP, Li T, Deeks JJ, editor(s).Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Hohmann 2018

Hohmann E, Glatt V, Tetsworth K, Cote M.Arthroscopic partial meniscectomy versus physical therapy for degenerative meniscus lesions: how robust is the current evidence? A critical systematic review and qualitative synthesis. *Journal of Arthroscopic and Related Surgery* 2018;**34**(9):2699-708.

Howell 2014

Howell R, Kumar NS, Patel N, Tom J.Degenerative meniscus: pathogenesis, diagnosis, and treatment options. *World Journal of Orthopedics* 2014;**5**(5):597-602.

Ike 1992

Ike RW, Arnold WJ, Rothschild E, Shaw HL, the Tidal Irrigation Cooperating Group.Tidal irrigation versus conservative medical management in patients with osteoarthritis of the knee: a prospective, randomized study. *Journal of Rheumatology* 1992;**19**:772-9.

Juhl 2012

Juhl C, Lund H, Roos EM, Zhang W, Christensen R.A hierarchy of patient-reported outcomes for meta-analysis of knee osteoarthritis trials: empirical evidence from a survey of high impact journals. *Arthritis* 2012;**2012**:136245.

Jüni 2015

Jüni P, Hari R, Rutjes AW, Fischer R, Silletta MG, Reichenbach S, et al.Intra-articular corticosteroid for knee osteoarthritis. *Cochrane Database of Systematic Reviews* 2015, Issue 10. Art. No: CD005328. [DOI: 10.1002/14651858.CD005328.pub3]

Kellgren 1957

Kellgren JH, Lawrence JS.Radiological assessment of osteo-arthrosis. *Annals of Rheumatic Disease* 1957;**16**:494–502.

Kellgren 2000

Kellgren JH, Lawrence JS.Radiological assessment of osteo-arthrosis. *Annals of Rheumatic Disease* 2000;**16**(4):494-502.

Khan 2014

Khan M, Evaniew N, Bedi A, Ayeni O, Bhandari M.Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. *CMAJ : Canadian Medical Association Journal* 2014;**186**:1057–64.

Lamplot 2016

Lamplot JD, Brophy RH.The role for arthroscopic partial meniscectomy in knees with degenerative changes: a systematic review. *Bone and Joint Journal* 2016;**98-B**:934-8.

Lee 2018

Lee DY, Park YJ, Kim HJ, Nam DC, Park JS, Song SY, et al.Arthroscopic meniscal surgery versus conservative management in patients aged 40 years and older: a meta-analysis. *Archives of Orthopaedic and Trauma Surgery* 2018;**138**(12):1731-9.

Leopoldino 2019

Leopoldino AO, Machado GC, Ferreira PH, Pinheiro MB, Day R, McLachlan AJ, et al.Paracetamol versus placebo for knee and hip osteoarthritis. *Cochrane Database of Systematic Reviews* 2019, Issue 2. Art. No: CD013273. [DOI: 10.1002/14651858.CD013273]

Li 2020

Li J, Zhu W, Gao X, Li X.Comparison of arthroscopic partial meniscectomy to physical therapy following degenerative meniscus tears: a systematic review and meta-analysis. *BioMed Research International* 2020;**2020**:1709415.

Lohmander 2019

Lohmander LS, Jarvinen TL.The importance of getting it right the first time. *Osteoarthritis and Cartilage* 2019;**27**(10):1405-7.

Mahir 2016

Mahir L, Belhaj K, Zahi S, Azanmasso H, Lmidmani F, El Fatimi A.Impact of knee osteoarthritis on the quality of life. *Annals of Physical and Rehabilitation Medicine* 2016;**59s**:e159.

McAlindon 2014

McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al.OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis and Cartilage* 2014;**22**:363-88.

Monk 2017

Monk AP, Garfjeld Roberts P, Palmer AJ, Bayliss LE, Beard DJ, Hopewell S, et al.The urgent need for evidence in arthroscopic meniscal surgery. *American Journal of Sports Medicine* 2017;**45**(4):965-73.

Mounsey 2009

Mounsey A, Ewigman B.Arthroscopic surgery for knee osteoarthritis? Just say no. *Journal of Family Practice* 2009;**58**(3):143-5.

NICE 2014

National Institute for Health and Care Excellence.Osteoarthritis: Care and management. London, UK: NICE 2014.

Niu 2011

Niu NN, Losina E, Martin SD, Wright J, Solomon DH, Katz JN.Development and preliminary validation of a meniscal symptom index. *Arthritis Care and Research* 2011;**63**(2):208-15.

Outerbridge 1961

Outerbridge RE.The etiology of chondromalacia patellae. *Journal of Bone and Joint Surgery. British Volume* 1961;**43**:752-7.

Page 2021

Page MJ, Higgins JP, Sterne JA.Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Pihl 2020

Pihl K, Ensor J, Peat G, Englund M, Lohmander S, Jørgensen U, et al.Wild-goose chase - no predictable patient subgroups who benefit from meniscal surgery: patient-reported outcomes of 641 patients 1 year after surgery. *British Journal of Sports Medicine* 2020;**54**(1):13-22.

Puljak 2017

Puljak L, Marin A, Vrdoljak D, Markotic F, Utrobicic A, Tugwell P.Celecoxib for osteoarthritis. *Cochrane Database of Systematic Reviews* 2017, Issue 5. Art. No: CD009865. [DOI: 10.1002/14651858.CD009865.pub2]

RACGP 2018

Royal Australian College of General Practitioners.Guideline for the management of knee and hip osteoarthritis. 2nd edition. East Melbourne, Australia: RACGP 2018.

Reichenbach 2010

Reichenbach S, Rutjes AW, Nüesch E, Trelle S, Jüni P.Joint lavage for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews* 2010, Issue 5. Art. No: CD007320. [DOI: 10.1002/14651858.CD007320.pub2]

Review Manager 2020 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager 5.4 (RevMan 5).Version 5.4. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2020.

Roemer 2017

Roemer FW, Kwoh CK, Hannon MJ, Hunter DJ, Eckstein F, Grago J, et al.Partial meniscectomy is associated with increased risk of incident radiographic osteoarthritis and worsening cartilage damage in the following year. *European Radiology* 2017;**27**:404-13.

Schünemann 2021a

Schünemann HJ, Vist GE, Higgins JP, Santesso N, Deeks JJ, Glasziou P, et al.Chapter 15: Interpreting results and drawing conclusions. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Schünemann 2021b

Schünemann HJ, Higgins JP, Vist GE, Glasziou P, Akl EA, Skoetz N, et al.Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

Shin 2012

Shin CS, Lee JH.Arthroscopic treatment for osteoarthritic knee. *Knee Surgery & Related Research* 2012;**24**(4):187-92.

Siemieniuk 2017

Siemieniuk RA, Harris IA, Agoritsas T, Poolman RW, Brignardello-Petersen R, de Velde SV, et al.Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. *BMJ* 2017;**257**:j1982.

Sihvonen 2018

Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H, et al.Arthroscopic partial meniscectomy versus placebo surgery for a degenerative meniscus tear: a 2-year follow-up of the randomised controlled trial. *Annals of the Rheumatic Diseases* 2018;**77**:188-95.

Steadman 2007

Steadman JR, Ramappa AJ, Maxwell RB, Briggs KK.An arthroscopic treatment regimen for osteoarthritis of the knee. *Arthroscopy* 2007;**23**(9):948-55.

Stensrud 2015

Stensrud S, Risberg MA, Roos EM.Effect of exercise therapy compared with arthroscopic surgery on knee muscle strength and functional performance in middle-aged patients with degenerative meniscus tears: a 3-mo follow-up of a randomized controlled trial. *American Journal of Physical Medicine and Rehabilitation* 2015;**94**(6):460–73. [NCT01002794]

Thorlund 2015

Thorlund JB, Juhl CB, Roos EM, Lohmander LS.Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms. *British Journal of Sports Medicine* 2015;**49**:1229-35.

Van de Graaf 2016

Van de Graaf VA, Wolterbeek N, Mutsaerts EL, Scholtes VA, Saris DB, De Gast A, et al.Arthroscopic partial meniscectomy or conservative treatment for nonobstructive meniscal tears: a systematic review and meta-analysis of randomized controlled trials. *Arthroscopy* 2016;**32**:1855–65.

Wai 2002

Wai E, Kreder H, Williams J.Arthroscopic debridement of the knee for osteoarthritis in patients fifty years of age or older: utilization and outcomes in the Province of Ontario. *Journal of Bone and Joint Surgery. American Volume* 2002;**84**:17-22.

Winter 2017

Winter AR, Collins JE, Katz JN.The likelihood of total knee arthroplasty following arthroscopic surgery for osteoarthritis: a systematic review. *BMC Musculoskeletal Disorders* 2017;**18**:408. [DOI: 10.1186/s12891-017-1765-0]

Zelen 1981

Zelen M.Alternatives to classic randomized trials. *Surgical Clinics of North America* 1981;**61**:1425-32.

Zhang 2009

Zhang W, Doherty M, Peat G, Bierma-Zeinstra MA, Arden NK, Bresnihan B, et al.EULAR evidence based recommendations for the diagnosis of knee osteoarthritis. *Annals of the Rheumatic Diseases* 2009;**69**(3):483-9.

Zhang 2010

Zhang Y, Jordan JM.Epidemiology of osteoarthritis. *Clinics in Geriatric Medicine* 2010;**26**(3):355-69.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Chang 1993

Study characteristics	
Methods	**Study design:** multicentre, parallel-group, two-arm randomised controlled trial **Setting:** the Rheumatology-Orthopedic Knee Clinic of the Northwestern Medical Faculty Foundation and the Division of Rheumatology of the Lutheran General Medical Group, Illinois, USA **Trial time period:** not reported **Interventions:** arthroscopic surgery versus non-arthroscopic joint lavage **Sample size calculations:** authors did not describe how the sample size was estimated **Analysis:** intention-to-treat analysis
Participants	**Number of participants** • Number screened: > 200 (110 ineligible) • Number of participants at enrolment: 90 (45 underwent arthroscopic surgery outside the study) • Number randomised: 34: 19 in the arthroscopy group and 15 in the lavage group • Number included in analyses: 32 participants, 18 in arthroscopy group, 14 in lavage group were included in the 3-month and 12-month analysis. Two participants (1 from arthroscopy group, 1 from lavage group) dropped out before treatment was given due to intercurrent medical problems and were excluded from the analysis. **Inclusion criteria** • Persistent knee pain for longer than 3 months, despite conservative medical and rehabilitation management, which restricted work, athletic, or self-care activities to an extent unacceptable to the patient • Weightbearing knee radiographs showing Kellgren-Lawrence (KL) classification grade 1, 2, or 3 changes • Age > 20 years • Willingness to attend follow-up visits at 3 and 12 months • Willingness to give written informed consent In participants with bilateral disease, the more symptomatic knee was designated the study knee.

Chang 1993 *(Continued)*

Exclusion criteria

- Knee surgery within 6 months of study entry
- Total knee replacement
- Any concurrent illness which would influence functional assessment of the knee or preclude arthroscopic surgery (e.g. severe intermittent claudication or cardiac disease)
- KL grade 4 changes or radiographs, as determined by one of the authors

Baseline characteristics

Arthroscopic surgery group (N = 18 included in analyses)

- Mean (SD) age: 61 (11) years
- % female: 72
- Mean (SD) duration of knee pain (months): 51 (51)
- Kellgren X-ray classification (%): Class I - 22, Class II - 28, Class III - 50
- Functional class (%): Class I - 5, Class II - 67, Class III - 28
- Walk distance (%): > 4 blocks - 44, 2 to 4 blocks - 11, < 2 blocks - 44
- Using assistive devise to walk (%): 50
- Mean (SD) initial active ROM (degrees): 114 (21)
- Knee joint tenderness measured on a 4-point ordinal scale, defined by the American College of Rheumatology (ACR) score of 3 or 4 (%): 29
- Knee joint swelling measured on a 4-point ordinal scale, defined by the ACR score of 3 or 4 (%): 41
- Mean (SD) initial Arthritis Impact Measurement Scale (AIMS) scores: pain 6.5 (2.0), physical activity 6.9 (2.0), physical function 2.3 (1.6), social activity 4.4 (2.2), depression 2.8 (2.2), anxiety 4.2 (2.4)
- Mean (SD) 50-foot walk time (seconds): 14.9 (4.3)
- Mean (SD) participant's global assessment (10 cm scale): 4.6 (2.6)
- Physician's global assessment score of 3 or 4 (%): 12

Joint lavage group (N = 14 included in analyses)

- Mean (SD) age: 65 (13) years
- % female: 71
- Mean (SD) duration of knee pain (months): 53 (57)
- Kellgren X-ray class (%): Class I - 14, Class II - 36, Class III - 50
- Functional class (%): Class I - 7, Class II - 79, Class III - 14
- Walk distance (%): > 4 blocks - 50, 2 to 4 blocks - 14, < 2 blocks - 36
- Using assistive devise to walk (%): 43
- Mean (SD) initial active ROM (degrees): 111 (20)
- Knee joint tenderness measured on a 4-point ordinal scale, defined by the ACR score of 3 or 4 (%): 31
- Knee joint swelling measured on a 4-point ordinal scale, defined by the ACR score of 3 or 4 (%): 21
- Mean (SD) initial AIMS scores: pain 6.1 (2.1), physical activity 5.3 (2.1), physical function 1.7 (1.0), social activity 4.7 (2.6), depression 2.6 (2.0), anxiety 3.9 (2.4)
- Mean (SD) 50-foot walk time (seconds): 15.0 (4.7)
- Mean (SD) participant's global assessment (10 cm scale): 4.6 (2.5)
- Physician's global assessment score of 3 or 4 (%): 16

Pre-treatment group differences: the baseline demographic, clinical and functional characteristics were similar between the two groups except for the initial AIMS Physical Activity Score (a statistically significant difference between groups (P < 0.05) was identified).

Interventions	**Arthroscopic surgery group** Arthroscopic surgery plus physical therapy and analgesia. Arthroscopy was done under general anaesthesia. A diagnostic evaluation was performed, and the anatomic findings were recorded on a standardised form. Following this evaluation, any of the following interventions were performed under arthroscopic guidance: (1) debridement of torn meniscus and removal of meniscal and cruciate liga-

Chang 1993 *(Continued)*

ment fragments; (2) removal of proliferative synovium; and (3) excision of loose articular cartilage fragments. Osteochondral lesions were not drilled. All participants received continuous saline lavage during the procedure and were routinely instructed in partial weightbearing precautions to continue for 10 days following the procedure. If an osteochondral lesion was detected in a weight-bearing area, this period of protection was increased to 3 weeks. Prior to and following surgery, participants assigned to this group received only non-narcotic analgesia and physical therapy, consisting of strengthening and flexibility exercises and gait training.

Non-arthroscopic (closed-needle joint) lavage group

Closed-needle joint lavage plus physical therapy and analgesia. Participants assigned to this group received non-narcotic analgesia and physical therapy identical to the arthroscopy group. In addition, participants received a tidal knee lavage procedure which was chosen to offset the potentially strong placebo effect of a surgical procedure and to control for the effects on pain and disability of the lavage procedure that occur during the arthroscopic procedure. Tidal knee lavage was performed as described by Ike and colleagues (Ike 1992) under local anaesthesia. A total of 1 litre of saline was injected into and aspirated from the knee in aliquots of 40 cc to 120 cc, depending on the size of the knee capsule.

Post-intervention

Participants in both groups received non-narcotic analgesia and physical therapy.

Outcomes

Outcomes were measured at baseline and at 3 and 12 months of follow-up

Clinical parameters

- Active and passive range of knee motion in degrees measured using goniometry
- Knee joint swelling measured on a 4-point ordinal scale, defined by the American College of Rheumatology (ACR) glossary. Improvement defined as a decrease of at least 1 point on the scale.
- Knee joint tenderness measured on a 4-point ordinal scale, as defined by the ACR glossary. Improvement defined as a decrease of at least 1 point on the scale.

Pain and functional status measures

- Arthritis Impact Measurement Scales (AIMS) scales for pain, physical function, physical activity, social activity, depression, and anxiety. The AIMS scales are scored from 0 (best) to 10 (worst) in a self-administered patient questionnaire. Improvement in the pain score was defined as a decrease of at least 1 point from the baseline score.
- Observed functional status was assessed with the 50-foot walk time in seconds.

Global measures

- Patient's global assessment (overall well-being) measured on a 10 cm visual analogue scale (VAS) scored from 0 (best) to 10 (worst). Improvement from the participant's perspective was defined as a reduction of > 1 cm from the baseline VAS score. Probability of improvement was defined as the proportion of participants who improved according to this criterion.
- Physician's global assessment of disease activity in the knee was made using a 4-point ordinal scale, ranging from no disease to very severe disease. Improvement from the physician's perspective was defined as a decrease of at least 1 point on the scale. Probability of improvement was defined as the proportion of participants who improved according to this criterion.

Economic measures

- Costs of the arthroscopic surgery or the tidal knee lavage
- Costs of medications
- Costs of other medical services, including physician and therapist visits
- Indirect costs of OA of the knee, including employment status and the use of paid and unpaid aides to help with activities of daily living

Outcomes included in this review at 3 and 12 months

Chang 1993 *(Continued)*

- Pain measured on AIMS-P subscale (0 to 10, lower score = less pain)
- Function measured on AIMS-PF subscale (0 to 10, lower score = better function) (we multiplied the mean values by −1 so that higher scores = better function, as per *Cochrane Handbook* guidance, so direction was consistent across function scales)
- Participant-reported treatment success (patient's global assessment) measured on VAS

Notes

Funding: supported by grant 9040 from the Robert Wood Johnson Foundation, by MAC grant AR-30692 from the NIH (NIAMS), and by the Percy Surgical Research Trust of Lutheran General Hospital

Trial registration: not reported

Adverse events: unclear if measured; not reported

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: not reported

Withdrawals: 2/34 (6%), 1 (5%) from arthroscopy group and 1 (7%) from lavage group, dropped out after randomisation but before treatment commenced and were excluded from analysis at 3 and 12 months.

Missing data at 12 months imputed by trialists using 3-month outcomes : 7/32 (22%), 5 (28%) from arthroscopy group, 2 (14%) from lavage group.

- 5 participants (4 from arthroscopy group, 1 from lavage group) dropped out and did not receive the assessment at the 12-month follow-up. No reason reported.
- 2 participants from lavage group were not satisfied with non-operative management and received arthroscopic surgery after 3 months of follow-up.

Treatment non-adherence: 3/18 participants from arthroscopy group withdrew before surgery; one improved between enrolment and planned surgery so cancelled surgery, and two developed other medical illnesses that precluded surgery.

Data analysis: missing standard deviations for outcomes used in this review: contacted authors but no response received. We imputed SDs.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Used a modification of the pre-randomisation design described by Zelen (Zelen 1981). Eligible participants were asked if they would accept an arthroscopic procedure if it was offered. Subjects who answered 'yes' were randomly assigned to arthroscopy or lavage and then asked to accept the assigned therapy. The randomisation plan was stratified by study site. No description of sequence generation process provided
Allocation concealment (selection bias)	Unclear risk	Insufficient description of the method of concealment. Use of assignment 'envelopes' described but unclear if safeguards used (e.g. sealed, sequentially numbered, opaque)
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding of participants and investigators was not done
Blinding of outcome assessor Self-reported outcomes	High risk	Blinding of participants was not done; hence, there was a risk of bias in the measurement of subjective outcomes of pain, physical function, physical activity, social activity, depression, and anxiety (AIMS subscales) and participants' global assessment

Chang 1993 *(Continued)*

Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	Blinding of outcome assessors was done. Participants were asked not to disclose their treatment assignment to assessors and to cover actual, or potential, arthroscopy scars
Incomplete outcome data (attrition bias) All outcomes	Low risk	Withdrawals (n = 2) were balanced in number across groups and for similar reasons. Missing data at 12 months were imputed using appropriate methods
Selective reporting (reporting bias)	Unclear risk	Trial registration not done and study protocol not available. Insufficient information to judge high or low risk
Other bias	Low risk	No other biases apparent

Gauffin 2014

Study characteristics	
Methods	**Study design:** single centre, parallel-group, two-arm, randomised, controlled trial **Setting:** orthopaedic department at the Linkoping University hospital, Sweden **Trial time period:** participants were enrolled between 4 March 2010 and 5 April 2012 **Interventions:** arthroscopic surgery plus unsupervised exercise versus unsupervised exercise **Sample size calculations:** the article did not describe how the sample size was estimated in this study. **Analysis:** intention-to-treat analysis was used; for those crossing over to the other group, an 'as-treated' analysis was performed.
Participants	**Number of participants** • Number screened: 179 (24 ineligible, 5 declined participation) • Number of participants at enrolment: 150 • Number randomised: 150: 75 in the surgery group and 75 in the non-surgery group • Number included in analyses: 123 participants were included in the 3-month analysis; 130 participants were included in the 12-month analysis; 119 participants were included in the 3-year analysis; 102 participants were included in the 5-year analysis (66 in arthroscopy group, 36 in exercise group). 9 participants in the arthroscopy group did not receive surgery. By 5 years, 19 participants crossed over from the exercise group to the arthroscopy group. 4 participants were excluded from the 5-year analysis because they underwent arthroscopic surgery in their index knee after the 12-month follow-up (n = 1 in arthroscopy group who hadn't had surgery as allocated and n = 3 in the exercise group). **Inclusion criteria** • Age 45 to 64 years with suspected meniscal injury • Symptom duration more than 3 months, including sudden pain onset and may include catching or locking of the joint (for < 2 seconds) • Standing X-ray with Ahlback grade of 0 (less than 50% reduction of the joint space, without consideration of possible osteophytes) • Had undergone prior physiotherapy • Could understand the Swedish language **Exclusion criteria** • Had a locked knee or joint • Locking for more than 2 seconds more often than once a week

Gauffin 2014 *(Continued)*

- Rheumatic or neurological disease
- Fibromyalgia
- Replacement of hip or knee joints
- Contraindication for day surgery at the current unit (BMI > 35 or a serious medical illness)

Baseline characteristics

Knee arthroscopy (N = 75)

- Mean (SD) age: 54 (5) years
- Number of men and women: M/F = 53/22
- Median (range) duration of knee pain (months): 7 (8)
- No. (%) expectations of treatment: 70 (97)
- Kellgren–Lawrence grade — no. (%): 0 - 37 (49); 1 - 34 (45); 2 - 4 (5)
- No. (%) sudden onset of pain: 45 (61)
- No. (%) daily joint catching: 45 (61)
- No. (%) joint locking for > 2 seconds: 18 (24)
- No. (%) mod to high physical activity level (PAS 4-6): 23 (32)
- Mean (95% CI) KOOS 5 subscales: Pain 55 (51-59); Symptoms 59 (55-62); Activities of Daily Living (ADL) 65 (61-69); Sports 29 (25-34); QOL 34 (30-37)
- Mean (95% CI) EQ-5D: Index 0.63 (0.57-0.68); VAS 62 (58-67)

Exercise therapy (N=75)

- Mean (SD) age: 54 (6) years
- Number of men and women: M/F = 56/19
- Median (range) duration of knee pain (months): 7 (7)
- No. (%) expectations of treatment: 67 (92)
- Kellgren–Lawrence grade — no. (%): 0 - 32 (43); 1 - 36 (48); 2 - 7 (9)
- No. (%) sudden onset of pain: 34 (47)
- No. (%) daily joint catching: 41 (56)
- No. (%) joint locking for > 2 seconds: 11 (15)
- No. (%) mod to high physical activity level (PAS 4-6): 23 (32)
- Mean (95% CI) KOOS 5 subscales: Pain 58 (54-62); Symptoms 62 (57-66); Activities of Daily Living (ADL) 68 (63-73); Sports 31 (26-37); QOL 35 (31-39)
- Mean (95% CI) EQ-5D: Index 0.62 (0.56-0.68); VAS 64 (59-69)

Pre-treatment group differences: there were no differences in the baseline characteristics between the two groups

Interventions	**Arthroscopic surgery** Arthroscopic surgery plus unsupervised exercise program. All operations were performed with full or local anaesthetics by an experienced arthroscopist at an independent daycare clinic. After the arthroscope was inserted in the joint and the joint was visually inspected, the surgeon judged, according to their experience, whether a meniscal resection or any other surgical treatment was indicated. After surgery, all participants were allowed immediate, full weight-bearing activity. The participants were advised to resume the exercise programme according to phase 1 for 1 week, and then switch to phase 2. Phase 1 program was for 1 week and consisted of a brisk walk 20-30 min, 10 x 2 sets of squats, pelvic lifts, pelvic lifts with ball between knees and extension of one knee, heel raise, wall squat, and standing on a pillow on one leg (30 sec x 2). Phase 2 was done twice a week for 3 months and consisted of 3 sets of all of the above-mentioned exercises. **Exercise** Unsupervised exercise program. At an independent clinic, five physiotherapists experienced in knee rehabilitation gave individual instructions for the exercise programme. The exercise programme aimed to increase muscle function and postural control and was done twice a week for 3 months, unsupervised.

Gauffin 2014 *(Continued)*

	Participants could exercise either at the gym or at home. Compliance was monitored with self-reported exercise diaries.
Outcomes	Outcomes were measured at baseline, at 3 and 12 months, 3 years and 5 years of follow-up. **Primary outcome** KOOS (Knee injury and Osteoarthritis Outcome Score) Pain subscale at 12 months, where scores range from 0 to 100 (100 indicates good knee function and less pain) **Secondary outcomes** • KOOS subscales (pain, symptoms, ADL, sports/recreation, QoL); scores range from 0 to 100 where higher scores indicate better knee function • Health-related quality of life using EQ-5D Index, a single index summarising responses in 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). EQ-5D VAS scores range from 0 (worst imaginable health state) to 100 (best imaginable health state) • Physical activity scale (PAS), a 6-point Likert scale ranging from "1: no physical activity" to "6: heavy physical activity several times a week" • Functional tests used at 3 months: squatting (possible without pain, possible with pain, not possible). SOLEC (standing on one leg with eyes closed test), 30-second chair stand test on one leg (maximum repetitions in 30 sec). • Symptom satisfaction scale, a 6-point Likert scale that ranged from “delighted” to “terrible" • Participant expectations about recovery, a 4-point Likert scale that ranged from "no recovery" to "full recovery" • Clinically important change in pain - improved (> 10-point higher KOOS Pain subscore at 5 years compared with baseline); stable (KOOS Pain subscore at 5 years within 10 points from baseline); deteriorated (> 10-point lower KOOS Pain subscore at 5 years compared with baseline) • Adverse events - electronic medical charts checked at 12 months and 3 years; and via questionnaire at 5 years • Radiographic changes at 5 years - radiographic OA defined as grade equal or greater than Kellgren-Lawrence grade 2 (definite osteophytes and possible joint space narrowing). Kellgren-Lawrence classification (grade 1 (doubtful): doubtful joint space narrowing and possible osteophytic lipping; grade 2 (minimal): definite osteophytes and possible joint space narrowing; grade 3 (moderate): moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends; grade 4 (severe): large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends). Deterioration in radiographic findings also assessed. **Outcomes included in this review at 3 and 12 months and at 5 years:** • Pain measured on KOOS Pain subscale (0 to 100; higher score = less pain) (we multiplied the mean values by –1 so that lower scores = less pain, as per *Cochrane Handbook* guidance) • Function measured on KOOS ADL subscale (0 to 100; higher score = better function) • Knee-specific health-related quality of life measured on KOOS QoL subscale (0 to 100; higher score = better QoL) • Generic health-related quality of life measured on EQ-5D Index (0 to 1, higher score = better QoL) • Participant-reported treatment success measured as improvement in the KOOS-Pain subscore of > 10 points at 5 years compared with baseline • Serious adverse events • Total adverse events • Progression of knee OA
Notes	**Funding:** no funding source reported **Trial registration:** NCT01288768 **Adverse events:** *Arthroscopic surgery*

Gauffin 2014 *(Continued)*

Serious adverse events:

No.(%): 3/66 (4.5%)

Nature of event: two repeat arthroscopies in the surgery group - one at 10 months and the other at 21 months after intervention. One participant died three years after surgery; it was not reported whether it was related to the intervention.

Other adverse events: none reported

Total adverse events:

No.(%): 3/66 (4.5%)

Exercise therapy

Serious adverse events: none reported

Other adverse events: none reported

Total adverse events: none reported

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: radiographic deterioration from baseline to the 5-year follow-up, assessed according to the Kellgren-Lawrence grade, occurred in 33/55 (60%) of the arthroscopy group and 10/27 (37%) of the exercise group.

Withdrawals: no withdrawals were reported in this study; however, 9/75 in the arthroscopy group and 18/75 in the exercise group did not complete the 3-month analysis; 5/75 in the arthroscopy group and 15/75 in the exercise group did not complete the 12-month analysis; 9/75 in the arthroscopy group and 39/75 in the exercise group did not complete the 5-year questionnaire and 20/75 in the arthroscopy group and 48/75 in the exercise group did not complete the 5-year weight-bearing radiographs. Reasons for lack of outcome data were not reported by group.

Data analysis: ITT data were used for 3- and 12-month follow-up but only as-treated data were reported at 5-year follow-up for outcomes: pain (Analysis 2.1), function (Analysis 2.2) and knee-specific and generic quality of life (Analysis 2.3; Analysis 2.4) and participant reported success (Analysis 2.5).

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No description of sequence generation process provided
Allocation concealment (selection bias)	Low risk	"The orthopaedic surgeon who enrolled and assessed participants was blinded to the allocation sequence. The allocations were placed in sequentially numbered, opaque, sealed envelopes in 15 blocks, block size 10"
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants were not blinded to group assignments. Blinding of physiotherapists was attempted but some participants revealed their group. Surgeons were not blinded
Blinding of outcome assessor Self-reported outcomes	High risk	Participants were not blinded to group assignments; thus, there was a risk of bias in the measurement of pain, function, knee-specific quality of life, generic quality of life
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	It is unclear whether outcome assessors checking electronic medical files for adverse events at 12 months were blind to group allocation, but we judged outcome measurement unlikely to be influenced by lack of blinding. Blinding of radiologist assessing X-rays not reported. Blinding of surgeon assessing X-

Gauffin 2014 *(Continued)*

		rays not done. Blinding of physiotherapists assessing functional tests (not used in our analysis) attempted but broken
Incomplete outcome data (attrition bias) All outcomes	High risk	9/75 in the arthroscopy group and 18/75 in the exercise group did not complete the 3-month analysis; 5/75 in the arthroscopy group and 15/75 in the exercise group did not complete the 12-month analysis; 8/74 in the arthroscopy group and 36/72 in the exercise group did not complete the 5-year analysis (4 participants underwent surgery in the affected knee and were excluded from the study after 1 year). In the 3-month assessment, 14 participants (per group information was not given) completed the questionnaire at 5 months and these data were excluded
Selective reporting (reporting bias)	High risk	Authors have not collected or reported Tegner activity scale and have collected but not reported KOOS total score (secondary outcomes in the trial registration form)
Other bias	Low risk	No other bias apparent

Herrlin 2007

Study characteristics	
Methods	**Study design:** single centre, parallel-group, two-arm, randomised, controlled trial **Setting:** Capio Artro Clinic, Stockholm Sports Trauma Research Center, Sweden **Trial time period:** participants were enrolled between June 2003 and April 2005 **Interventions:** arthroscopic surgery plus supervised exercise versus supervised exercise **Sample size calculations:** sample size calculations reported that 40 participants per group were needed to detect an average difference of 10 points in the KOOS with 80% power. P = 0.05 was considered to be statistically significant. **Analysis:** intention-to-treat
Participants	**Number of participants** • Number screened: 180 (80 declined participation, 3 ineligible) • Number of participants at enrolment: 97 • Number randomised: 96 (1 left the study after randomisation as his knee returned to normal function), 47 in the arthroscopy group and 49 in the exercise group • Number included in analyses: 92 at 24 months (46 in the arthroscopy group and 46 in the exercise group); 92 at 60 months (45 in the arthroscopy group and 47 in the exercise group) **Inclusion criteria** • Age 45 to 64 years • Knee pain without a trauma, daily or almost daily pain experienced during the last 2 to 6 months • Knee osteoarthritis grade 0 or 1 on weight-bearing knee radiographs according to Ahlbacks classification • Medial meniscal tear on MRI • Understanding of the Swedish language. **Exclusion criteria** • Traumatic meniscal injury • Neurological and rheumatic inflammatory diseases

Herrlin 2007 *(Continued)*

- Loose bodies, ligament injuries, osteochondral defects and tumours (MRI)
- Earlier knee surgery, prosthetic replacements of the hip or knee and fractures to the lower extremities less than 1 year old
- Contraindication to physical training

Baseline characteristics

Arthroscopy (N = 47)

- Mean (SD) age (men): 54 (4.6) years
- Mean (SD) age (women): 54 (5.6) years
- Number of men and women: M/F = 28/19
- Mean (SD) body-mass index (men): 27 (3.9)
- Mean (SD) body-mass index (women): 25 (3.9)
- Pain medication (no.) men - 5
- Pain medication (no.) women - 5
- Median (IQR 25-75) Tegner activity scale: 3 (0-6)
- Mean (SD) KOOS 5 subscales (data supplied by authors): pain 55.9 (18.7); symptoms 63.1 (17.4); ADL 66.4 (20.1); sports 26.1 (23.1); QoL 36.3 (18.8)
- Mean (SD) Lysholm total score: 60.9 (14.6)
- Mean (SD) VAS, movement: 5.8 (2.6)
- Mean (SD) VAS, rest: 2.4 (2.4)

Exercise (N = 49)

- Mean (SD) age (men): 55 (5.5) years
- Mean (SD) age (women): 59 (3.8) years
- Number of men and women: M/F = 27/16
- Mean (SD) body-mass index (men): 26 (3.3)
- Mean (SD) body-mass index (women): 25 (6.5)
- Pain medication (no.) men - 3
- Pain medication (no.) women - 3
- Median (IQR 25-75) Tegner activity scale: 3 (0-7)
- Mean (SD) KOOS 5 subscales (data supplied by authors): pain 62.6 (19.1); symptoms 67.0 (17.9); ADL 69.7 (21.2); sports 36.5 (27.3); QoL 37.3 (17.4)
- Mean (SD) Lysholm total score: 68.2 (15.7)
- Mean (SD) VAS, movement: 4.6 (2.7)
- Mean (SD) VAS, rest: 2.4 (2.58)

Pre-treatment group differences: there were no differences in baseline characteristics between the two groups

Interventions	**Arthroscopic surgery** Arthroscopic partial meniscectomy plus supervised exercise. Arthroscopy was performed on an outpatient basis by two experienced surgeons, majority under local anaesthesia. A 5.5 mm, 30, arthroscope was used with a pressure-controlled irrigation system. A standard operation protocol was used to document possible findings in cartilage, ligaments, synovium and the medial and lateral meniscus. Meniscal lesions were registered and changes in the articular cartilage were classified according to the Outerbridge classification: grade 0 = intact articular surfaces, grade I = softening of the surfaces, grade II = partial-thickness defects less than 1.5 cm, grade III = partial-thickness tears greater than 1.5 cm/fragmentation, grade IV = exposed bone. Twice a week during a period of 8 weeks, each participant followed a standardised exercise programme which consisted of exercises for improving muscle strength and endurance, muscle flexibility, balance and proprioception. The goal of the exercise programme was to reduce pain, restore full ROM and improve knee function. The participants were informed to exert the exercises with some strain but perform them almost pain-free and without having any negative influence in the affected knee at the fol-

Herrlin 2007 *(Continued)*

lowing day. If the participant could tolerate the exercises without any problems, he/she performed the exercises with increasing weights and higher resistance.

Exercise
Supervised exercise programme. Each participant followed a standardised exercise programme with the possibility for individual adaptation twice a week for a period of 8 weeks, as described above.

Outcomes

Outcomes were measured at baseline and at 2, 6, 12, 24 and 60 months' follow-up (the published papers reported at 2, 6, 24 and 60 months; author supplied 12-month data upon request).

Outcomes

- Knee injury and Osteoarthritis Outcome Scale (KOOS) comprising 5 subscales - pain, symptoms, ADL, sports/rec, QoL, where scores range from 0 to 100 where 100 indicates good knee function
- Lysholm Knee Scoring Scale, where scores range from 0 to 100 with higher scores indicating better knee function without symptoms
- Tegner Activity Scale, which covers activities in daily life and sports. Scores range from 0 to 100 with higher scores indicating better activity/ function
- Knee pain during rest and during weight bearing measured on 10-point Visual Analogue Scales (VAS). Scores range from 0 = no pain to 10 = maximal pain
- Degree of osteoarthritis based on radiographic assessment at 60 months' follow-up and classified according to Ahlback classification (grade 1: joint space narrowing (less than 3 mm); grade 2: joint space obliteration; grade 3: minor bone attrition (0-5 mm); grade 4: moderate bone attrition (5-10 mm); grade 5: severe bone attrition (more than 10 mm)). Radiographs taken in weight-bearing in antero-posterior projection, in lateral projection with the knee joint in semi flexion and skyline view of patella with knee flexed at 30 degrees. Progression of knee OA appeared to be defined post hoc of at least one grade progression on the Ahlback classification

Outcomes used in this review at 2 and 6 months, 2 and 5 years

- Pain measured on KOOS Pain subscale (0 to 100; higher score = less pain) (we multiplied the mean values by –1 so that lower scores = less pain, as per *Cochrane Handbook* guidance)
- Function measured on KOOS ADL subscale (0 to 100, higher score = better function)
- Knee-specific health-related quality of life measured on KOOS QoL subscale (0 to 100, higher score = better QoL)
- Progression of knee osteoarthritis
- Adverse events

Notes

Funding: not reported

Trial registration: not done

Adverse events:

Arthroscopic surgery

Serious adverse events: No. (%): 3/47 (6.4%)

Three participants had an additional arthroscopy between 13 and 40 months after the initial operation. Re-arthroscopy showed degenerative articular changes.

Other adverse events: 0/47 (0%)

Total adverse events: 3/47 (6.4%)

3 participants from the arthroscopy group had an additional arthroscopic procedure between 13 and 40 months following the original surgery.

Exercise

Serious adverse events: 13/49 (26.5%)

Herrlin 2007 *(Continued)*

13 participants presented with medial meniscal tears on arthroscopic examination and were treated with partial meniscectomy.

Other adverse events: 0/49 (0%)

Total adverse events: 13/49 (26.5%)

Cross-overs: 13 participants from the exercise group had an arthroscopy at an average of 6.5 months after the intervention.

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: 2/43 participants from the arthroscopy group and 2/45 participants from the exercise group who had radiographic examination at 60 months after the intervention had progression of knee osteoarthritis in the medial compartment. In three cases, progression was from grade 1 to grade 2 and in one case from grade 1 to grade 3 (the authors did not specify which treatment group the latter participant was from)

Withdrawals: none

Data analysis: KOOS Pain, KOOS ADL and KOOS QoL were extracted at 2, 6, 12 and 24 months based on data supplied by the author upon request; progression of knee osteoarthritis was extracted from published paper.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No description of sequence generation process provided
Allocation concealment (selection bias)	Unclear risk	There was no reporting of allocation concealment in this study
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding of participants and study personnel was not done
Blinding of outcome assessor Self-reported outcomes	High risk	As blinding of participants was not done, there is a risk of bias in the measurement of knee pain and function
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	No blinding of study personnel but the assessor-reported outcomes, adverse events and progression of knee OA based on radiographic examination, are unlikely to be influenced by lack of blinding
Incomplete outcome data (attrition bias) All outcomes	Low risk	All participants who completed baseline measurements completed follow-up assessments at 8 weeks and 6 months. 4 participants were lost to follow-up at 24 months (1 from arthroscopy group and 3 from exercise group) and 4 participants were lost to follow-up at 60 months (2 from arthroscopy group and 2 from exercise group)
Selective reporting (reporting bias)	High risk	Trial registration not done. 12-month outcomes were not reported in the trial publications
Other bias	Low risk	No other bias apparent

Katz 2013

Study characteristics	
Methods	**Study design:** multicentre, parallel-group, two-arm, randomised, controlled trial **Setting:** seven US tertiary referral centres - Brigham and Women's Hospital, Boston; Hospital for Special Surgery, New York; Cleveland Clinic, Cleveland; Vanderbilt University, Nashville; Mayo Clinic, Rochester; Rush University Medical Center, Chicago; Washington University, St. Louis, USA **Trial time period:** participants were enrolled between June 2008 and August 2011 **Interventions:** arthroscopic surgery plus supervised physical therapy versus supervised physical therapy **Sample size calculations:** the study was powered to detect a 10-point difference on the WOMAC functional scale between the operative and non-operative arms, The study team adopted a Type I error rate of 5% and power of 80% and the target sample size was set at 340 participants. The sample size calculation took into account two sources of sample degradation: losses to follow-up and cross-over from the assigned arm to the other arm prior to the primary outcome assessment at six months. The study was also powered for one pre-planned subgroup analysis in which participants with Kellgren-Lawrence (KL) Grade 3 (joint space narrowing) would be analysed in one subgroup and those with KL Grades 0 to 2 in the other. **Analysis:** outcome analysis only included participants who did not withdraw from the study.
Participants	**Number of participants** • Number screened: 14,430 (12,008 did not meet inclusion criteria, 1092 not screened by physician, 195 eligible but not referred, 784 declined to participate) • Number of participants at enrolment: 351 • Number randomised: 351: 174 in the arthroscopy group and 177 in the physical therapy group. 51 crossed over to surgery group at 6 months and 59 crossed over to surgery group at 12 months • Number included in analyses: 330: 161 in the arthroscopy group and 169 in the physical therapy group (at 6 months), and 320; 156 in the arthroscopy group and 164 in the physical therapy group (at 12 months) **Criteria for defining knee osteoarthritis with meniscal tear:** symptomatic participants 45 years of age or older with a meniscal tear and evidence of mild-to-moderate osteoarthritis on imaging **Inclusion criteria**. • Symptoms for at least four weeks, managed with one or more of: medications, activity limitations or PT • Symptoms consistent with torn meniscus (at least one of: clicking, catching, popping, giving way, pain with pivot or torque, pain that is episodic, pain that is acute and localised to one joint line) • Availability of knee radiograph and MRI • Evidence on knee MRI of osteophytes or full-thickness cartilage defect; or plain radiographic evidence of osteophytes or joint space narrowing • Evidence on knee MRI of a meniscal tear that extends to the surface of the meniscus • Willingness to undergo randomisation and ability to understand and sign an informed consent document **Exclusion criteria** • A chronically locked knee (e.g. participant cannot reduce locking; a clear-cut indication for arthroscopic partial meniscectomy) • Kellgren-Lawrence Grade 4 (far advanced OA) • Inflammatory arthritis or clinically symptomatic chondrocalcinosis • Injection with viscosupplementation in past four weeks in index knee • Contraindication to surgery or physical therapy

Katz 2013 *(Continued)*

- Bilateral symptomatic meniscal tears
- Prior surgery on same knee

Baseline characteristics

Arthroscopic Partial Meniscectomy (N = 161)

- Mean (SD) age: 59.0 (7.9) years
- Number of men and women: M/F = 71/90
- Treatment history: pharmacological and physical therapy
- Concurrent treatment: NSAIDs, acetaminophen, intra-articular glucocorticoids
- Mean (SD) body-mass index: 30.0 (6.1)
- Mean (SD) WOMAC physical-function score: 37.1 (17.9)
- Mean (SD) KOOS pain score: 46.0 (15.5)
- Mean (SD) Mental Health Index 5 score: 74.8 (12.9)
- Mean (SD) SF-36 physical-activity score: 44.3 (23.7)
- Kellgren–Lawrence grade — no. (%): 0- 34 (21.1); 1- 26 (16.1); 2- 37 (23.0); 3- 45 (28.0)

Physical Therapy (N = 169)

- Mean (SD) age: 57.8 (6.8 years)
- Number of men and women: M/F = 72/97
- Treatment history: pharmacological and physical therapy
- Concurrent treatment: NSAIDs, acetaminophen, intra-articular glucocorticoids
- Mean (SD) body-mass index: 30.0 (6.1)
- Mean (SD) WOMAC physical-function score: 37.5 (18.3)
- Mean (SD) KOOS pain score: 47.2 (16.4)
- Mean (SD) Mental Health Index 5 score: 74.0 (13.9)
- Mean (SD) SF-36 physical-activity score: 43.3 (23.3)
- Kellgren–Lawrence grade — no. (%): 0- 36 (21.3); 1- 35 (20.7); 2- 39 (23.1); 3- 39 (23.1)

Pre-treatment group differences: no differences were reported between the two groups.

Interventions	**Arthroscopic surgery** Arthroscopic partial meniscectomy plus physical therapy. The damaged meniscus was trimmed back to a stable rim. Loose fragments of cartilage and bone were removed without any penetration of the subchondral bone. Preoperative antibiotics were used routinely. Postoperatively, participants were allowed to bear weight as they were able. Bracing was not used. Participants were referred to a physical therapist for a postoperative standardized physical therapy program, as described below. **Exercise:** Supervised physical therapy provided to both groups. ***Phase I-Acute Phase*** (1-10 days post-op) 8 exercises, 12-15 repetitions, 1-2 sets of the following types of exercises: - *Decrease Inflammation*: Retrograde Massage, Cryotherapy E-Stim: NMES or IFC - *Manual Therapy:* Joint Mobilization Soft Tissue Mobilization Stretching LE Muscles - *Open Chain Exercises*: Quad Sets SAQ/LAQ/HS Curls Hip-4 way - *Closed Chain Exercises*: Bicycle, Elliptical, Treadmill, Leg Press, Balance/Proprioception ***Phase II-Subacute Phase*** (10 days - 4 weeks post-op) Participant must meet 3 of the 4 criteria: Knee A/PROM 0>=115 degrees, moderate to minimal effusion, knee Pain = 3/5)

Katz 2013 *(Continued)*

8 exercises, 12-15 repetitions, 1-2 sets of the following types of exercises:

- *Decrease Inflammation*: Retrograde Massage Cryotherapy E-Stim: NMES or IFC
- *Manual Therapy*: Joint Mobilization Soft Tissue Mobilization Stretching LE Muscles
- *Open Chain Exercises*: Add more Concentric/Eccentric Hip/Knee progressive resistive exercises, ROM
- *Closed Chain Exercises*: Resisted terminal knee extension, modified mini squats, step up/down progressions, toe raises, functional and agility training

Phase III-Advanced Activity Phase (4-7 weeks post-op)

8 exercises, 12-15 repetitions, 1-2 sets of the following types of exercises:

- *Continued stretching program*
- *Continued PRE therapeutic exercises program*
- *Emphasis on closed chain program with progression to dynamic single leg stance, plyometrics, running, and sport specificity training*

In both the arthroscopic partial meniscectomy plus physical therapy and physical therapy alone groups, participants were permitted to receive acetaminophen and non-steroidal anti-inflammatory agents as needed, and intra-articular injections of glucocorticoids were permitted over the course of the trial.

Outcomes

Outcomes were measured at baseline and at 3 and 6 months and every 6 months thereafter up to 5 years follow-up (protocol states outcomes measured at baseline, 3, 6, 12, 18 and 24 months, but data in published papers report outcomes at baseline, 6 and 12 months and up to 5 years. Authors did not respond to requests for data).

Primary outcome

Physical function on the physical-function scale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) assessed at 6 months. Scores range from 0 to 100, with higher scores indicating worse physical function

Secondary outcomes

- Pain and symptoms measured on the Knee Injury and Osteoarthritis Outcome Scale (KOOS) Pain and Symptoms subscales respectively. Scores range from 0 to 100, with higher scores indicating more pain (data on KOOS symptoms subscale not reported)
- SF-36 Physical Activity subscale. Scores range from 0 to 100, with higher scores indicating greater physical activity
- Treatment success - improvement in the WOMAC physical-function score of at least 8 points (defined as a clinically relevant difference specified a priori)
- Mental health status measured on MHI-5 (5-Item Mental Health Index). Scores range from 0 to 100, with higher scores indicating better mental health
- Physical examination - quadriceps strength, passive and active knee extension and flexion, gait disturbance, balance, proprioception and functional limitation (measured at baseline and 3 months only) (data not reported)
- Generic health-related quality of life measured on the EQ-5D (data not reported)
- Adverse events
- Knee joint replacement
- Medical comorbidity measured on the Self Administered Comorbidity Questionnaire (data not reported)
- Health care utilisation - hospitalisations, medications, assistive devices, tests, procedures and visits to physicians, physical therapists and other providers, direct non-medical costs (incl. transportation to appointments, time and expenses involved in receiving care), indirect costs (lost wages, time off work) (data not reported)

Outcomes included in this review at 3, 6 and 24 months and 5 years

- Pain measured on KOOS Pain subscale (0 to 100, lower score = less pain)

Katz 2013 *(Continued)*

- Function measured on WOMAC Physical Function scale (0 to 100, lower score = better function) (we multiplied the mean values by –1 so that higher scores = better function, as per *Cochrane Handbook* guidance, so direction was consistent across function scales)
- Participant-reported success measured as improvement in the WOMAC Physical Function score of at least 8 points
- Knee surgery - self-report in questionnaire and medical record review
- Serious adverse events
- Total adverse events

Notes

Funding: this study was funded by grants (R01AR055557, K24AR057827, and P60AR047782) from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health.

Trial registration: ClinicalTrials.gov number, NCT00597012 (METEOR study)

Adverse Events

Arthroscopic partial meniscectomy (APM):

Serious adverse events - no. (%)

- Cardiovascular - 2 (1.15%) (Pulmonary embolism resulting in death- 1; Acute myocardial infarction -1)
- Vascular disorders (Hypoxemia) - 1 (0.57%)
- Knee replacement - 5 (2.8%)

Total serious adverse events - 8/164 (4.9%)

Other adverse events - no. (%)

- Cardiovascular - 6 (3.45%) (Deep vein thrombosis - 2; syncope - 1; skin - 2; other -1)
- Musculoskeletal - 7 (4.02%) (Pain from fall or other trauma - 2; tendonitis - 3; rupture of baker's cyst - 1; knee pain - 1)
- Pain in the back, hip, or foot - 2 (1.15%)

Total other adverse events - 15 (8.62%)

Total ALL adverse events - 23/164 (14.0%)

Physical therapy (PT):

Serious adverse events - no. (%)

- Cardiovascular- 2 (1.13%) (Stroke - 1; Sudden death - 1)
- Knee replacement - 3 (1.7%)

Total serious adverse events - 5/109 (4.59%)

Other adverse events - no.(%)

- Cardiovascular - 3 (1.69%) (Atrial fibrillation - 1; skin -1; other -1)
- Musculoskeletal - 6 (3.39%) (Pain from fall or other trauma - 4; knee bursitis - 1; knee pain - 1)
- Pain in the back, hip, or foot - 4 (2.26%)

Total other adverse events - 13 (7.34%)

Total ALL adverse events - 18/109 (16.5%)

There was no group differences in the frequency of adverse events.

Knee surgery (replacement or osteotomy) (12 months' follow-up):

Arthroscopic Partial Meniscectomy:

Katz 2013 *(Continued)*

No. of participants = 5 (2.8%)

Physical Therapy:

No. of participants = 3 (1.7%)

Knee surgery (replacement or osteotomy) (5 years' follow-up):

Randomised to and receiving Arthroscopic Partial Meniscectomy:

No. of participants = 16/164 (9.8%)

Cross-overs from Physical Therapy to Arthroscopic Partial Meniscectomy:

No. of participants = 7/68 (10.3%)

Randomised to and receiving Physical Therapy:

No. of participants = 2/109 (1.8%)

Progression of knee OA: not reported

Withdrawals: 7/174 from APM group and 4/177 from physical therapy group at 6 months and 9/174 from the APM group and 7/177 from physical therapy group at 12 months

Report of study results: Attempts to obtain KOOS Pain and WOMAC PF 5-year follow-up data from authors were unsuccessful. For total knee replacement, adverse events and serious adverse events, we reported 109 in the PT group (i.e. those who did not cross over to APM).

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomisation was done using a secure program on the trial website" in blocks of varying size within each site, stratified by the extent of osteoarthritis at baseline (Kellgren-Lawrence grade 0–2 (no joint space narrowing) versus Kellgren-Lawrence grade 3 (< 50% joint space narrowing). Probably low risk
Allocation concealment (selection bias)	Low risk	Randomisation was performed by a research coordinator in real time using a secure website, thus ensuring concealment until the time of allocation to treatment group
Blinding of participants and personnel (performance bias) All outcomes	High risk	"Our study was not blinded, since our investigative group did not consider a sham comparison group feasible"
Blinding of outcome assessor Self-reported outcomes	High risk	As participants were aware of their treatment, assessment of self-reported outcomes including WOMAC and KOOS were at risk of detection bias
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	High risk	Radiographs of the knee were assessed by surgeons (who were aware of the treatment assignment) as well as musculoskeletal radiologists, whose knowledge of treatment assignment is not reported. KL grading and assessment of radiographs are subject to bias as blinding of outcome assessors was not done
Incomplete outcome data (attrition bias) All outcomes	Low risk	The proportion of loss to follow-up and reasons for incomplete outcome data are similar between groups

Katz 2013 *(Continued)*

Selective reporting (reporting bias)	High risk	SF-36 5-item mental health index and EQ-5D were measured, according to the protocol paper, but were not listed as outcomes in the trial registration form. No results data for these outcomes are available
Other bias	Unclear risk	33.8% (60/177) of participants assigned to physical therapy crossed over to arthroscopy within 6 months of randomisation, but their outcome data for pain, function and treatment success was included in the analysis for the physical therapy group potentially underestimating any effect of surgery

Kirkley 2008

Study characteristics	
Methods	**Study design:** single centre, parallel-group, two-arm, randomised, controlled trial **Setting:** Fowler Kennedy Sport Medicine Clinic, University of Western Ontario, London, Ontario, Canada **Trial time period:** participants were enrolled between January 1999 and August 2007 **Interventions:** arthroscopic surgery plus optimised physical and medical therapy, home exercises and arthritis education versus optimised physical and medical therapy, home exercises and arthritis education **Sample size calculations:** assignment of 186 participants to treatment would provide 80% statistical power to detect a 200-point difference in the WOMAC score between the two treatment groups with allowance of 15% of participants whose data cannot be evaluated **Analysis:** intention-to-treat analysis was conducted
Participants	**Number of participants** • Number screened: 277 (58 ineligible, 31 declined participation) • Number of participants at enrolment: 188 • Number randomised: 188: 94 in the arthroscopy group and 94 in the physical and medical therapy group. 6 participants in the surgery group declined surgery • Number included in analyses: 168: 88 in the arthroscopy group and 80 in the physical and medical therapy group (at 24 months) **Criteria for defining knee osteoarthritis:** participants with idiopathic or secondary osteoarthritis of the knee with grade 2, 3, or 4 radiographic severity, as defined by the modified Kellgren–Lawrence classification **Inclusion criteria** • 18 years of age or older • Idiopathic or secondary osteoarthritis of the knee • Grade 2, 3, or 4 radiographic severity, as defined by the modified Kellgren–Lawrence classification **Exclusion criteria** • Large meniscal tears (“bucket handle” tears) as detected by clinical examination or by MRI • Inflammatory or post-infectious arthritis • Previous arthroscopic treatment for knee osteoarthritis • More than 5 degrees of varus or valgus deformity • Previous major knee trauma, Kellgren–Lawrence grade 4 osteoarthritis in two compartments (the medial or lateral compartments of the tibiofemoral joint or the patellofemoral compartment) in persons over 60 years of age

Kirkley 2008 *(Continued)*

- Intra-articular glucocorticoid injection within the previous 3 months
- A major neurologic deficit
- Serious medical illness (life expectancy of less than 2 years or high intra-operative risk)
- Pregnancy
- Inability to provide informed consent or who were deemed unlikely to comply with follow-up

Baseline characteristics

Arthroscopic Surgery (N = 92)

- Mean (SD) age: 58.6 (10.2) years
- Number of men and women: M/F = 38/52
- Mean (SD) body-mass index: 31.6 (6.7)
- Mean (SD) WOMAC total score 1187 (483); pain dimension 239 (105); stiffness dimension 117 (50); physical-function score: 830 (355)
- Mean (SD) standard-gamble utility score 0.79 (0.23)
- Mean (SD) duration of osteoarthritis symptoms in study knee (months) 47.1 (69.4)
- Mean (SD) SF-36 physical activity score: 33.8 (7.6)
- Kellgren–Lawrence grade — no. (%): 2- 42 (46); 3- 45 (49); 4-5 (5)
- No. (%) symptoms of catching or locking: 48 (52)
- No.(%) joint effusion: 56 (61)
- No. (%) positive McMurray test: 1 (1)
- No. (%) pain with forced flexion: 62 (67)
- No. (%) tenderness at the tibiofemoral joint line: 81 (88)
- No. (%) magnetic resonance imaging performed: 15 (16)

Physical and Medical Therapy (N = 86)

- Mean (SD) age: 60.6 (9.9) years
- Number of men and women: M/F = 28/58
- Mean (SD) body-mass index: 30.2 (6.3)
- Mean (SD) WOMAC total score 1043 (542); pain dimension 214 (122); stiffness dimension 103 (48); physical-function score: 726 (397)
- Mean (SD) duration of osteoarthritis symptoms in study knee (months) 40.1 (72.6)
- Mean (SD) SF-36 physical activity score: 33.9 (8.6)
- Mean (SD) standard-gamble utility score 0.81 (0.20)
- Kellgren–Lawrence grade — no. (%): 2- 36 (42) ; 3- 46 (53); 4- 4 (5)
- No. (%) symptoms of catching or locking: 38 (44)
- No.(%) joint effusion: 53 (62)
- No. (%) positive McMurray test: 1 (1)
- No. (%) pain with forced flexion: 56 (65)
- No. (%) tenderness at the tibiofemoral joint line: 77 (90)
- No. (%) magnetic resonance imaging performed: 10 (12)

Pre-treatment group differences: participants in the surgery group had slightly higher total WOMAC scores compared to those in the physical and medical therapy group. Other baseline characteristics were similar across the two groups.

Interventions	**Arthroscopic surgery:** Arthroscopic surgery plus optimised physical and medical therapy, home exercises and arthritis education. Arthroscopic debridement was performed under general anaesthesia. Irrigation of the medial, lateral, and patellofemoral joint compartments was done using 1 litre of saline. One or more of the following procedures was done - synovectomy; debridement; or excision of degenerative tears of the menisci, fragments of articular cartilage, or chondral flaps and osteophytes. Optimised physical and medical therapy was initiated within 7 days after surgery and followed an identical program in both groups.

Kirkley 2008 *(Continued)*

Exercise:

Physical and medical therapy, home exercises and arthritis education provided to both groups. Physical therapy was provided for 1 hour once a week for 12 consecutive weeks. Information regarding a home exercise program that emphasised range-of-motion and strengthening exercises was provided to all participants. These exercises were done twice daily and once on the day of a scheduled physical-therapy session. After 12 weeks of supervised activity, participants continued an unsupervised exercise program at home for the duration of the study. The participants received additional education from attendance at local Arthritis Society workshops, from a copy of The Arthritis Helpbook that was provided to them, and from an educational videotape.

The medical therapy for both groups involved the administration of intra-articular injection of hyaluronic acid, oral glucosamine and step-wise use of acetaminophen and non-steroidal anti-inflammatory drugs.

Outcomes

Outcomes were measured at baseline and at 3, 6, 12, 18 and 24 months of follow-up

Primary outcome

The total Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) score (range, 0 to 2400; higher scores indicate more severe symptoms and poorer physical function) at 2 years of follow-up.

Secondary outcomes

- WOMAC - Pain, Stiffness and Physical Function subscales scores (range from 0 to 2400; higher scores indicate more severe pain and stiffness and poorer knee function)
- Short Form-36 (SF-36) Physical Component Summary score (range 0 to 100; higher scores indicate better quality of life)
- McMaster-Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) assessing symptoms and functional status (scores range from 0 to 500; higher scores indicate greater disability)
- Arthritis Self-Efficacy Scale (ASES) assessing perceived ability to cope with the consequences of arthritis (scores range from 10 to 100; higher scores indicate greater self-efficacy)
- Health-related quality of life was assessed by the standard-gamble utility technique; scores can range from 0.0 (death) to 1.0 (perfect health)

Outcomes used in this review at 3, 6 and 24 months

- Pain measured on WOMAC-Pain subscale (0 to 2400, lower score = less pain)
- Function measured using WOMAC Physical Function subscale (0 to 2400, lower score = better function) (we multiplied the mean values by –1 so that higher scores = better function, as per *Cochrane Handbook* guidance, so direction was consistent across function scales)
- Generic health-related quality of life measured using the standard-gamble utility score (score range 0 = death to 1 = perfect health)

Notes

Funding: this study was sponsored by the Fowler Kennedy Sport Medicine Clinic and supported by the Canadian Institutes of Health Research.

Trial registration: Clinical trials number NCT00158431

Adverse events: unclear if measured; not reported

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: not reported

Withdrawals: 6/94 in the arthroscopy group and 14/94 in the physical and medical therapy group

Risk of bias

Bias	Authors' judgement	Support for judgement

Kirkley 2008 *(Continued)*

Random sequence generation (selection bias)	Low risk	Participants were randomly assigned using a computer-generated schedule.
Allocation concealment (selection bias)	Low risk	To minimise the risk of predicting the treatment assignment of the next eligible participant, randomisation was performed in permuted blocks of two or four with random variation of the blocking number.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding of study staff was done - "To preserve blinding, each patient wore a neoprene sleeve over the knee so that the study nurse could not identify a surgical scar". Blinding of participants was not done.
Blinding of outcome assessor Self-reported outcomes	High risk	As blinding of participants was not done, there could be bias in the assessment of pain, stiffness, disability, quality of life and functional status.
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	"At each visit, the patients were evaluated by a nurse who was unaware of the treatment assignment."
Incomplete outcome data (attrition bias) All outcomes	High risk	6/94 (6%) (withdrawal of consent - 2; withdrawal from study - 1; death - 1; loss to follow-up - 2) in the arthroscopy group and 14/94 (15%) (withdrawal of consent - 8; loss to follow-up - 6) in the physical and medical therapy group did not complete the study
Selective reporting (reporting bias)	Low risk	All outcomes listed in the trial registration and protocol have data reported in the results publication.
Other bias	Low risk	No other bias apparent

Kise 2016

Study characteristics	
Methods	**Study design:** multicenter, parallel-group, two-arm, randomised controlled trial **Setting:** orthopaedic departments at two public hospitals and two physiotherapy clinics in Norway **Trial time period:** participants were enrolled between October 2009 and September 2012 **Interventions:** arthroscopic surgery versus supervised exercise **Sample size calculations:** the sample size calculation was based on the change in $KOOS_4$ (defined as the average score for 4/5 KOOS subscale scores) from baseline to 2-year follow-up. To detect a 10-point difference with a standard deviation of 15, with a level of power of 90%, level of significance of 0.05, and an estimated 15% dropout rate at two years, 56 participants were required in each group. To factor in a 20% cross-over rate, 140 participants were recruited. **Analysis:** intention-to-treat
Participants	**Number of participants** • Number of participants screened: 341 (115 not eligible) • Number of participants at enrolment: 226 (85 declined participation, 1 ineligible) • Number randomised: 140: 70 in the arthroscopic partial meniscectomy group and 70 in the supervised exercise therapy group

Kise 2016 *(Continued)*

- Number included in analyses: 129 at 3 months (64 in the surgery group and 65 in exercise group), 129 at 12 months (66 in the surgery group and 63 in the exercise group), 126 at 24 months (64 in the surgery group and 62 in the exercise group), 120 at 5 years (62 in the surgery group and 58 in the exercise group). Fourteen out of 70 participants (20%) in the exercise group crossed over to the arthroscopic group and were analysed in the meniscectomy group in the 'as-treated analysis'

Definition of degenerative meniscal tear: degenerative meniscal tear was defined as an intramenisical linear magnetic resonance imaging signal penetrating one or both surfaces of the meniscus.

Inclusion criteria

- Age 33 to 60 years
- Knee pain for more than 2 months without a significant trauma
- Medial meniscal tear on MRI
- Eligible for arthroscopic partial meniscectomy
- Grade 0 to 2 radiographic severity (specification after Kellgren-Lawrence)

Exclusion criteria

- Those requiring acute trauma surgeries, including those treated as acute cases in the emergency room
- Ligament injuries
- Tumours (MRI)
- Pain or other musculoskeletal comorbidities severely affecting lower extremity muscle function overriding the symptoms from the knee
- Grade 3 or 4 radiographic severity (specification after Kellgren-Lawrence)
- Comorbidities excluding physical activities and exercise
- Previous knee surgery within two years
- Not able to speak or read Norwegian, drug abuse or mental problems

Baseline characteristics

Arthroscopic partial meniscectomy (N = 70)

- Mean (SD) age: 48.9 (6.1) years
- No. (%) men: 43 (61)
- No. (%) right knee: 41 (59)
- Mean (SD) body-mass index: 26.0 (3.7)
- No. (%) smokers: 10 (14.3)
- No. (%) use analgesics daily: 3 (4.2)
- Severity of radiographic osteoarthritis in knee - Kellgren-Lawrence classification: 0 - 51 (73), 1 - 16 (23), 2 - 3 (4), 3 - 0
- No. (%): meniscal degeneration grade 1-2: 6 (9), grade 3a-3b: 64 (91)
- No. (%): meniscal extrusion: 35 (50)
- Mean (SD) pain duration (months): 12.0 (15.7)
- Mean (SD) knee function VAS: 63.8 (18.9)
- Mean (SD) KOOS scores: $KOOS_4$ - 59.6 (13.8); pain - 67.6 (14.9); symptoms - 77.4 (14.6); daily activities 79.6 (16.1); function in sport and recreation 47.8 (23.4); knee-related quality of life 45.6 (15.5)
- Mean (SD) SF-36: Physical component 47.4 (6.1), Mental component 56.0 (6.3)
- Mean (SD) Peak torque extension 163.1 (53.2), Total work extension 790.8 (254.8), Peak torque flexion 88.5 (25.7), Total work flexion 492.9 (158.7)
- Mean (SD) One leg hop test (cm) 83.2 (35.5)
- Mean (SD) 6-metre timed hop test (sec) 2.7 (1.2)
- Mean (SD) Knee bends 30 sec test (No) 29.3 (10.6)

Supervised exercise therapy (N = 70)

- Mean (SD) age: 50.2 (6.2) years

Kise 2016 *(Continued)*

- No. (%) men: 43 (61)
- No. (%) right knee: 41 (59)
- Mean (SD) body-mass index: 26.4 (4.3)
- No. (%) smokers: 3 (4.2)
- No. (%) use analgesics daily: 3 (4.2)
- Severity of radiographic osteoarthritis in knee - Kellgren-Lawrence classification: 0 - 49 (70), 1 - 18 (26), 2 - 2 (3), 3 - 1 (1)
- No. (%): meniscal degeneration grade 1-2: 6 (9)
- No. (%): meniscal degeneration grade 3a-3b: 63 (91)
- No. (%): meniscal extrusion: 45 (65)
- Mean (SD) pain duration (months): 17.3 (21.5)
- Mean (SD) knee function VAS: 57.9 (21.5)
- Mean (SD) KOOS scores: $KOOS_4$ - 54.3 (18.2); pain - 63.4 (20.8); symptoms - 69.8 (16.7); daily activities 75.0 (21.5); function in sport and recreation 44.0 (25.8); knee-related quality of life 40.0 (17.5)
- Mean (SD) SF36: Physical component 45.4 (8.4), Mental component 55.0 (9.2)
- Mean (SD) Peak torque extension 157.5 (48.7), Total work extension 772.9 (245.1), Peak torque flexion 81.9 (27.2), Total work flexion 448.3 (187.8)
- Mean (SD) One leg hop test (cm) 76.6 (32.8)
- Mean (SD) 6-metre timed hop test (sec) 3.1 (1.7)
- Mean (SD) Knee bends 30 sec test (No) 28.2 (10.6)

Pre-treatment group differences: there were no obvious differences in baseline characteristics between the groups.

Interventions

Arthroscopic surgery

Arthroscopic partial meniscectomy. Six orthopaedic surgeons with at least 10 years of clinical experience performed the operations. Surgery was performed with the participant under general anaesthesia, with or without thigh tourniquet, antibiotic prophylaxis, or antithrombotic prophylaxis. Arthroscopes with 30 degree optics and standard arthroscopic instruments were used. Ringer acetate was used for lavage. Normal procedure involved two portals: anteromedial and anterolateral. Additional injuries (ligaments, cartilage) preceded systematic probing of both menisci, and finally, all unstable meniscal tissue was resected. The participants were discharged from hospital on the day of surgery and were advised to use two crutches post-operatively until gait normalised and no swelling or discomfort occurred during weight bearing. Before hospital discharge, the participants were given written and oral instructions for simple home exercises, aimed at regaining knee range of motion and reducing swelling. They were encouraged to perform the exercises two to four times daily.

Exercise

The supervised exercise therapy programme consisted of progressive neuromuscular and strength exercises over 12 weeks, performed during a minimum of two and a maximum of three sessions each week (24 to 36 sessions). Each session lasted approximately 60 to 80 minutes. About 20 minutes was spent warming up and cooling down on a stationary cycle, 20 to 30 minutes was spent on neuromuscular exercise, and 20 to 30 minutes was spent on strength training.

Outcomes

Outcomes were measured at baseline and at 3, 12 and 24 months and 5 years of follow-up.

Primary outcomes

- $KOOS_4$ defined as the average score for 4/5 KOOS subscale scores on pain, symptoms, function in sport and recreation, and knee-related quality of life at 24 months. Subscale scores are calculated separately and transformed to a scale from 0 to 100 where 0 = worst, 100 = best
- Isokinetic muscle strength at 3 months (including peak torque flexion, peak torque extension, total work flexion and total work extension) for knee extensors and flexors
- Incident and enlarging marginal tibiofemoral osteophyte at 5 years (co-primary outcome pre-specified in trial registration record)

Kise 2016 *(Continued)*

Secondary outcomes

- Five KOOS subscales - pain, symptoms, ADL, sports/rec, QoL. Subscale scores are calculated separately and transformed to a scale from 0 to 100 where 0 = worst, 100 = best
- SF-36 Physical and Mental Component Summaries (scores range from 0 = worst to 100 = best)
- 3 performance tests to evaluate lower extremity function:
  a. the one leg hop test for distance (measuring length in centimetres, higher score indicative of better function)
  b. the 6-metre timed hop test (measuring time in seconds, lower time indicative of better function),
  c. the knee bends test (measuring maximum number in 30 seconds, higher score = better function).
- Adverse events - any situations where participants sought health care
- Serious adverse events - death, cardiovascular or gastrointestinal events, deep vein thrombosis, pulmonary embolism, systemic or local infection
- Progression of osteophytes and tibiofemoral joint space narrowing at 5 years assessed by the Osteoarthritis Research Society International (OARSI) atlas - scored separately for the medial and lateral tibiofemoral compartments on a 0 (normal) to 3 (severe changes) scale (half-grades were used when progression had occurred but without achieving a full grade; participants undergoing subsequent tibial osteotomy were given joint space narrowing score of 3)
- Total radiographic progression score at 5 years - sum of medial and lateral compartment joint space narrowing and osteophyte score
- Severity of knee osteoarthritis at 5 years - using the modified Kellgren-Lawrence classification from 0 (normal) to 4 (severe), including a grade with definite osteophyte only (grade 2/ost)
- Radiographic knee osteoarthritis at 5 years - assessed as Grade 2 (definite osteophyte and possible joint space narrowing) or above using Kellgren-Lawrence classification (grade 1 (doubtful): doubtful joint space narrowing and possible osteophytic lipping; grade 2 (minimal): definite osteophytes and possible joint space narrowing; grade 3 (moderate): moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends; grade 4 (severe): large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends) in knees graded as 0 or 1 at baseline. Those receiving total knee arthroplasty also categorised as having radiographic knee osteoarthritis
- Symptomatic knee osteoarthritis at 5 years - participants reporting knee pain at least weekly (question P1 from KOOS Pain subscale) in addition to incident radiographic knee osteoarthritis
- Knee surgery (replacement or osteotomy)

Outcomes included in this review at 3 and 24 months and 5 years

- Pain measured on KOOS Pain subscale (0 to 100, higher score = less pain) (we multiplied the mean values by −1 so that lower scores = less pain, as per *Cochrane Handbook* guidance, so direction was consistent across pain scales)
- Function measured on KOOS ADL subscale(0 to 100, higher score = better)
- Knee-specific health-related quality of life measured on KOOS QoL subscale(0 to 100, higher score = better)
- Generic health-related quality of life measured on SF-36 Mental Component Summary (0 to 100, higher score = better) (24 months only)
- Serious adverse events
- Total adverse events
- Progression of knee osteoarthritis
- Knee surgery (replacement or osteotomy)

Notes

Funding: this study was funded by Sophies Minde Ortopedi AS, Swedish Rheumatism Association, Swedish Scientific Council, Region of Southern Denmark, Danish Rheumatism Association, and the Health Region of South-East Norway.

Trial Registration: www.clinicaltrials.gov (NCT01002794)

Adverse events:

Arthroscopic partial meniscectomy

Kise 2016 *(Continued)*

Serious adverse events

- Serious knee-related adverse events: 1 participant in the surgery group received a total knee arthroplasty 34 months after the index surgery. 2 participants (one allocated to surgery and one cross-over) received an osteotomy 4 to 6 months after the index procedure. 4 participants (three allocated to surgery and one cross-over) underwent another partial meniscectomy at 6, 12, 15 and 36 months after the index operation.
- Other serious adverse events: none reported

Other adverse events

- Pain, swelling, instability, stiffness, decreased range of motion in the index knee - 16 (23%)
- Similar symptoms in the contralateral knee - 10 (14%)
- Total other adverse events - 26 (37%)

Total ALL adverse events: 31/64 (48.4%)

Supervised exercise therapy

Serious adverse events

- Serious knee-related adverse events: none reported
- Other serious adverse events: none reported

Other adverse events

- Pain, swelling, instability, stiffness,decreased range of motion in the index knee - 16 (23%)
- Similar symptoms in the contralateral knee - 15 (21%)
- Total other adverse events - 31 (44%)

Total ALL adverse events: 31/60 (51.7%)

Knee surgery (replacement or osteotomy) (5 year follow-up): 1 participant in the surgery group received a total knee arthroplasty 34 months after the index surgery.

2 participants (one allocated to surgery and one cross-over) received an osteotomy 4 to 6 months after the index procedure

Progression of knee OA: 10/58 (17.2%) participants from the exercise group and 13/62 (21.0%) participants from the arthroscopy group had radiographic knee OA consistent with grade 2 or more on the Kellgren-Lawrence classification at five years.

Withdrawals: 6/70 in the arthroscopic partial meniscectomy group and 8/70 in the supervised exercise therapy group

Cross-overs: in the final report (Berg 2020, secondary publication of Kise 2016), 14/70 crossed over from exercise to arthroscopy group in contrast to earlier report (Kise 2016) which reported 13/70 cross-overs.

Report of study results: 3-month study findings reported in Stensrud 2015 (secondary publication of Kise 2016).

Data analysis: the authors supplied outcome data for knee-related quality of life measured on KOOS QoL subscale (0 to 100, higher score = better) upon request

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation sequence was used.

Kise 2016 *(Continued)*

Allocation concealment (selection bias)	Low risk	The allocations were kept in sequentially-numbered, opaque envelopes.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants were not blinded to the group assignments; however, strength and function test assessors were blinded.
Blinding of outcome assessor Self-reported outcomes	High risk	Participants were not blinded to group assignments, thus, there was a risk of bias in the measurement of pain, function and knee-related quality of life.
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	The outcome assessors were blinded to group allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	8/70 in the exercise group and 6/70 in the arthroscopic group did not have outcome data (questionnaires not returned).
Selective reporting (reporting bias)	Unclear risk	One outcome (incident and enlarging marginal tibiofemoral osteophytes at 5 years) was pre-specified as a co-primary outcome but moved to secondary outcome and redefined as progression of both osteophytes and tibiofemoral joint space narrowing separately at 5 years. The rationale for the change was provided (to describe radiographic changes in knee osteoarthritis development) and the change was made before 5-year data collection and analysis. Additional secondary outcomes that were not pre-specified were reported (e.g. total radiographic progression score, severity of knee OA using modified Kellgren-Lawrence classification, adverse events, knee surgery). Although the paper did not publish the KOOS summary data for each treatment group, the authors provided them upon request and they were used in the analysis.
Other bias	Low risk	No other bias apparent

Merchan 1993

Study characteristics	
Methods	**Study design:** single centre, parallel-group, two-arm, open-label, randomised controlled trial **Setting:** participants recruited from outpatient clinic of an orthopaedic hospital in Spain **Trial time period:** January 1988 to December 1990 **Interventions:** arthroscopic surgery plus physiotherapy versus non-steroidal anti-inflammatory drugs (NSAIDs) plus physiotherapy **Sample size calculations:** a priori sample size calculation not described **Analysis:** intention-to-treat
Participants	**Number of participants** Number of participants screened: not reported Number of participants at enrolment: not reported

Merchan 1993 *(Continued)*

Number randomised: 80: 40 to arthroscopic surgery and physiotherapy group and 40 to non-steroidal anti-inflammatory drugs and physiotherapy group

Number included in the analysis: data for 35/40 (87.5%) for the arthroscopic surgery and physiotherapy group and 38/40 (95%) for the non-steroidal antiinflammatory drugs and physiotherapy group were available at the 1-, 2- and 3-year follow-ups

Inclusion criteria

- Age > 50 years
- Painful limited degenerative osteoarthritis of the femorotibial joint
- Limited degenerative process (minimal joint space narrowing and formation of small osteophytes) on pre-operative radiographs evaluated according to Ahlback 1968

Exclusion criteria

- Duration of pain > 6 months
- Body weight > 85 kg in men and > 70 kg in women
- History of previous surgery of the affected knee
- Participants with an appreciable instability or an angular deformity of > 15 degrees on standing antero-posterior radiographs for varus and valgus femorotibial angulation
- Participants with femoro-patellar joint involvement

Baseline characteristics

Arthroscopic surgery and physiotherapy group (N = 35)

Mean (range) age: 57.1 (50 to 63) years

No. of men/women: 7/28

Mean Inital Knee Score (IKS) calculated using the modified Hospital for Special Surgery Knee Rating Score: 26.85

Mean Final Knee Score (FKS) calculated using the modified Hospital for Special Surgery Knee Rating Score: 37.00

Mean Knee Score Difference (KSD = FKS-IKS): 10.14

Non-steroidal anti-inflammatory drugs and physiotherapy group (N = 38)

Mean (range) age: 56.9 (50 to 65) years

No. of men/women: 13/25

Mean IKS: 29.86

Mean FKS: 32.76

Mean KSD: 2.89

Pre-treatment group differences: the initial knee score (IKS), age and sex did not show statistical differences between the two groups.

Interventions	**Arthroscopic surgery** Arthroscopic surgery and physiotherapy. The surgical technique included debridement of synovial tissue; removal of degenerative menisci, osteophytes, and loose bodies; and limited debridement of cartilage defects. Osseous debridement to bleeding bone (abrasion arthroplasty) was not performed. The mean operative time was 50 minutes (range 35 to 70). Post-operatively, a compression bandage was used with early exercises, motion, and weight bearing as tolerated. The operative findings of articular cartilage changes were graded according to Outerbridge (Outerbridge 1961). Superficial fibrillation was present in 8 participants, fragmentation of < 1.3cm^2 in 16, fragmentation of > 1.3cm^2 in 7, and eburnation to subchondral bone in 4. There were meniscal tears in 31 of the 35 knees with a ratio of medial

Merchan 1993 *(Continued)*

	to lateral of 4 to 1. The most common tear was a flap of the posterior horn of the medial meniscus. Osteophytes were excised from the intercondylar notch in 6 knees, loose bodies were removed in 7, and 4 were noted to have chondrocalcinosis. Physiotherapy was practiced for 4 weeks after surgery. The physiotherapy regimen included quadriceps exercises and knee flexion exercises immediately postoperatively. **Non-steroidal anti-inflammatory drugs** Non-steroidal anti-inflammatory drugs and physiotherapy. The non-operative treatment consisted of non-steroidal anti-inflammatory drugs and a decrease in the intensity of the activities of daily living for a pain-free knee. Physiotherapy was practiced as in the operative group (i.e. quadriceps and knee flexion exercises for 4 weeks). This non-operative group had no further courses of treatment during the follow-up period.
Outcomes	The mean follow-up time was 25 months (range 12 to 36) in the arthroscopic surgery and physiotherapy group and 23 months (range 12 to 36) in the non-steroidal anti-inflammatory drugs and physiotherapy group. **Outcomes** • Modified Hospital for Special Surgery Knee Rating Score (mHSSKRS) (0 to 100 points, higher score = better). mHSSKRS has two subscores: subjective score = 60 points (includes pain = 35, instability = 10, walking aids = 5, walking distance = 10); and objective score = 40 points (includes extension = 10, flexion = 20, effusion = 10). (Only group means were reported. No measures of variability reported) • Treatment success defined as an increase in the post-treatment knee rating score of at least 10 points. A failure was defined as a knee score decreasing or failing to increase by 10 points. • Adverse events **Outcomes used in this review at mean of 23 to 25 months** • Participant-reported treatment success (at least 10 point increase in post-treatment mHSSKRS) • Serious adverse events • Total adverse events Measures of variability for function outcomes using mHSSKRS were not reported so could not be used in this review.
Notes	**Funding:** Fundacion Caja de Madrid **Trial registration:** not done **Adverse events:** *Arthroscopic surgery and physiotherapy group* **Serious adverse events:** Death: 5/40 (12.5%) participants died after randomisation and were excluded from follow-up. No further information provided. Deep vein thrombosis: 2/40 (5%) **Other adverse events:** Superficial infection: 1/40 (2.5%) Haemarthrosis: 1/40 (2.5%) Total ALL adverse events: 9/40 (22.5%) *Non-steroidal anti-inflammatory drugs and physiotherapy group* **Serious adverse events:**

Merchan 1993 *(Continued)*

Death: 2/40 (5%) participants died after randomisation and were excluded from follow-up. No further information provided.

No other serious adverse events reported

Other adverse events: none reported

Total ALL adverse events: 2/40 (5%)

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: not reported

Withdrawals: 5/40 in the arthroscopic surgery and physiotherapy group and 2/40 in the non-steroidal anti-inflammatory drugs and physiotherapy group due to death

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No description of sequence generation process provided. There were some baseline differences in outcome measures between the treatment groups.
Allocation concealment (selection bias)	Unclear risk	Allocation performed by "pulling consecutively numbered envelopes that had previously been randomly placed on a bulletin board". Unclear if sealed or opaque
Blinding of participants and personnel (performance bias) All outcomes	High risk	No information provided. Probably not done
Blinding of outcome assessor Self-reported outcomes	High risk	No information provided. Probably not done. Measurement of mHSSKRS subjective score likely to be influenced by lack of blinding
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	High risk	No information provided. Probably not done. Measurement of mHSSKRS objective score and treatment success likely to be influenced by lack of blinding. Adverse events unlikely to be influenced by lack of blinding
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Imbalance in withdrawals across groups: 12.5% (5/40) due to death in the arthroscopic surgery plus physiotherapy group vs. 5% (2/40) due to death in the NSAIDs plus physiotherapy group. The cause of deaths in both groups were not reported.
Selective reporting (reporting bias)	Unclear risk	Trial registration not done and protocol not available. Insufficient information to judge high or low risk
Other bias	Low risk	No other bias apparent

Moseley 1996

Study characteristics

Methods	**Study design:** single centre, three-arm, double-blind, randomised, placebo-controlled trial **Setting:** Houston Veterans Administration (VA) Medical Center, Texas, USA

Moseley 1996 *(Continued)*

Trial time period: June 1992

Interventions: arthroscopic debridement versus arthroscopic lavage versus placebo surgery

Sample size calculations: a priori sample size calculation was not done

Analysis: statistical analysis not done

Participants

Number of participants

Number of participants screened for eligibility: not reported

Number of participants enrolled: 10

Number randomised: 10 participants: 5 participants were randomised to the placebo group, 3 were randomised to the arthroscopic lavage group and 2 to the arthroscopic debridement group

Number included in analysis: a statistical analysis was not performed because of the small number of subjects and responses. However, data were available from all 10 participants at 2 and 6 weeks and from 9 participants at 3 and 6 months and reported as means.

Inclusion criteria

- Symptomatic osteoarthritis of the knee in spite of a minimum of 6 months of non-operative treatment including non-steroidal anti-inflammatory medication
- At least moderate knee pain (~ 4 on a 0 to 10 scale) on average over a week's time
- Age under 70
- No medical problems that placed the participant at significant risk for complications from a general anaesthetic (as determined by the general medicine consult service at the Houston VA Medical Center)

Exclusion criteria

Not reported

Baseline characteristics

Arthroscopic debridement (N = 2)

Mean intensity of worst knee pain on a scale where 1 = no pain, 10 = severe pain: 9.0

Mean average intensity of knee pain on a scale where 1 = no pain, 10 = severe pain: 7.0

Mean intensity of today's pain on a scale where 1 = no pain, 10 = severe pain: 6.0

Number of days this week with knee pain: 6

Average knee extension (negative values refer to flexion contracture): -12.0

Average knee flexion: 111

Average knee crepitus on a scale: 0 = none, 3 = severe: 1.5

Average knee effusion on a scale: 0 = none, 3 = severe: 0

Average global knee tenderness on a scale: 0 = none, 9 = severe: 5.0

50-foot walk in seconds: 12.0

Arthroscopic Lavage (N = 3)

Mean intensity of worst knee pain on a scale where 1 = no pain, 10 = severe pain: 8.5

Mean average intensity of knee pain on a scale where 1 = no pain, 10 = severe pain: 7.5

Mean intensity of today's pain on a scale where 1 = no pain, 10 = severe pain: 5.5

Moseley 1996 *(Continued)*

	Number of days this week with knee pain: 6 Average knee extension (negative values refer to flexion contracture): -5.0 Average knee flexion: 105 Average knee crepitus on a scale: 0 = none, 3 = severe: 1.5 Average knee effusion on a scale: 0 = none, 3 = severe: 1.0 Average global knee tenderness on a scale: 0 = none, 9 = severe: 4.0 50-foot walk in seconds: 11.0 *Placebo surgery* (N = 5) Mean intensity of worst knee pain on a scale where 1 = no pain, 10 = severe pain: 8.4 Mean average intensity of knee pain on a scale where 1 = no pain, 10 = severe pain: 6.8 Mean intensity of today's pain on a scale where 1 = no pain, 10 = severe pain: 6.8 Number of days this week with knee pain: 5.6 Average knee extension (negative values refer to flexion contracture): -5.2 Average knee flexion: 125 Average knee crepitus on a scale: 0 = none, 3 = severe: 1.0 Average knee effusion on a scale: 0 = none, 3 = severe: 0 Average global knee tenderness on a scale: 0 = none, 9 = severe: 2.8 50-foot walk in seconds: 11.5 **Pre-treatment group differences:** there were no obvious differences in baseline characteristics between the three groups except for greater flexion contracture in the debridement group compared to the other two groups.
Interventions	The arthroscopic debridement and arthroscopic lavage groups both received a standard general endotracheal anaesthetic that was routinely used at the medical centre. The placebo arthroscopy group received a lesser anaesthetic, an intravenous tranquilliser (benzodiazepine, droperidol, or both) in conjunction with an opioid (fentanyl or sufentanil), without having placement of an endotracheal tube. This combination of drugs, along with the local anaesthetic injected at the stab wound sites, sedated the participant and rendered them insensitive to pain. The placebo participants breathed spontaneously throughout the procedure, and they inhaled oxygen supplemented by a nasal cannula. End-tidal carbon dioxide was continuously monitored throughout the procedure. Using this type of anaesthesia for the placebo group minimised the potential complications from induction of general anaesthesia and placement of an endotracheal tube. Once anaesthetised, all participants had their knees examined and then prepared and draped in the usual manner. All participants had a gram of cephalosporin antibiotic administered intravenously as prophylaxis against infection. Bupivacaine (0.25%) with epinephrine was then injected into the skin where the arthroscopy portals were to be made. **Arthroscopic debridement** The arthroscopic debridement participants had three stab wounds made and an arthroscope inserted in the inferolateral portal; an inflow cannula was inserted in the superomedial portal, and the various operating instruments were inserted from the inferomedial portal. The knee was distended with sterile saline from the inflow cannula, and a constant flow of fluid was lavaged through the knee as is typical for arthroscopic surgery. A minimum of 10 litres of fluid was lavaged through the knee. After a diagnostic arthroscopic examination, the arthroscopic instruments were used to shave the rough articular cartilage (chondroplasty), remove loose debris, trim torn or degenerated meniscal fragments, and correct

Moseley 1996 *(Continued)*

any other soft tissue abnormalities that could interfere with the mechanical function of the knee. However, no abrasion arthroplasty or removal of bone spurs was performed. At the end of the procedure, the instruments were removed, the portals were closed with absorbable suture, and a sterile compression dressing was applied. The average time for the surgery was 45 minutes.

Arthroscopic lavage

The arthroscopic lavage participants had a procedure identical to that of the debridement participants, except that no operating instruments were used to remove or trim the various parts of the knee. A diagnostic arthroscopic examination was performed, and a minimum of 10 litres of fluid was lavaged through the knee. A minimum of 30 minutes was spent performing the lavage procedure, and the total time spent in the operating room was approximately 1 hour.

Placebo surgery

Participants undergoing placebo arthroscopy were prepared, draped, examined and injected with local anaesthetic in the same manner as the other two groups. Three stab wounds were made in the skin with a scalpel, but no instruments of any kind were placed into the knee. The knee was manipulated, instruments were requested and passed, saline was splashed, and a standard arthroscopic debridement was simulated as closely as possible in the event the participant was not totally unaware during the event. A minimum of 30 minutes was spent performing the placebo surgery, and the typical time spent in the operating room was 1 hour.

Participants in all three groups spent approximately an hour in the operating theatre. All participants were taken to the recovery room and treated the same as any participant having any arthroscopic procedure. Once the participant returned to the orthopaedic ward from the recovery room, he/she was observed until stable enough to be sent home. All participants were discharged from the hospital the afternoon of surgery or the next morning with an oral narcotic analgesic (typically acetaminophen with codeine) for use as required. Before discharge, participants were fitted with crutches and instructed to discontinue using them as soon as they could walk comfortably without a limp. At the first post-operative visit, participants were instructed to resume their pre-operative anti-inflammatory medications and to resume walking and other activities of daily living as soon as their symptoms would allow.

Outcomes

Outcomes were measured at 2 and 6 weeks, and 3 and 6 months.

Study outcomes

- Pain (5 domains: intensity of worst knee pain; average intensity of knee pain; intensity of today's pain; pain relief experienced after surgery; number of days this week with knee pain). Domains 1 through 3 were each measured on a 10-point scale where 1 = no pain, 10 = severe pain. Domain 4 was measured on a 10-point scale where 1 = no relief, 10 = complete relief.
- Mobility (activity) was measured but the details of the measurement scale and the findings were not reported in this paper
- General well-being was measured but the details of the measurement scale and the findings were not reported in this paper
- Satisfaction with surgery, measured using two questions (would you recommend the surgery to your family or friends?; do you feel the operation was worthwhile?) each using a 5-point Likert scale (strongly agree to strongly disagree)
- 50-foot walk, time in seconds
- Knee extension and flexion measured via physical examination
- Knee crepitus and effusion measured on a 4-point scale: 0 = none, 3 = severe
- Knee tenderness measured on a 10-point scale: 0 = none, 9 = severe

The subjective data (pain, mobility, general well-being, satisfaction with surgery) were collected from a questionnaire designed specifically for this study (not included in trial report). The questionnaire was reported to be based on the AIMS-2, the SF-36, the Wisconsin Brief Pain Questionnaire and the Knee Society's Knee Rating Scale.

Outcomes used in this review at 3 and 6 months

- Pain - average intensity of knee pain, measured on a 10-point scale (lower score = less pain)

Moseley 1996 *(Continued)*

- Participant reported success - satisfaction with surgery at 6 months, measured as number of participants reporting 'strongly agree' or 'slightly agree' for item 'do you feel the operation was worthwhile?'

Notes

Funding: not reported

Trial registration: not reported

Adverse events: unclear if measured; not reported

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: not reported

Withdrawals: 1/10 in total. 1/2 (50%) from the arthroscopic debridement group

Analysis: we included arthroscopic debridement versus placebo and we excluded the lavage treatment arm from this review.

Data imputations: SDs for pain were not reported at 3 and 6 months (and no baseline values were reported). We used SD values from Sihvonen 2013 for pain at 3 and 6 months for this study in Analysis 2.1 as Sihvonen and colleagues also used a pain numerical rating scale. For satisfaction with surgery, we assumed dropouts (one in arthroscopic debridement group) were unsatisfied and used the number randomised as the denominator.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Insufficient information about sequence generation process given, other than stating it was randomised. "An unbalanced randomization scheme was used to determine if physician and patient blinding could truly be maintained in the placebo surgery group."
Allocation concealment (selection bias)	Unclear risk	Sealed 'randomisation envelopes' were used, which were opened in the operating theatre, to reveal which procedure the participant was to receive. Insufficient information provided about whether appropriate safeguards were used (e.g. opaque, sequentially-numbered envelopes)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants and study personnel (other than the operating surgeon) were blinded to the treatment assignment group. Identical preparation, stab wounds and post-operative care were performed in all groups so it is unlikely that participants could guess their group assignment. "All postoperative care was performed by orthopaedic residents, nurses and other personnel who were blinded to the type of treatment that the patient received."
Blinding of outcome assessor Self-reported outcomes	Low risk	Blinding of participants was done; low risk of bias in the measurement of pain and satisfaction with surgery
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	Knee replacement not measured
Incomplete outcome data (attrition bias) All outcomes	High risk	1/2 (50%) participants withdrew from the arthroscopic debridement group and 0/5 withdrew from the placebo surgery group. The reason for withdrawal was not reported.
Selective reporting (reporting bias)	High risk	Trial registration not done and protocol not available. Data on mobility and general well-being were reported as being collected in the methods but results

Moseley 1996 *(Continued)*

		not reported. Pain measured but outcome data not adequately reported (no measure of variance)
Other bias	Low risk	No other bias apparent

Moseley 2002

Study characteristics

Methods	**Study design:** single centre, three-arm, randomised, placebo-controlled trial **Setting:** Houston Veterans Affairs Medical Center, Houston, Texas, USA **Trial time period:** participants were enrolled from October 1995 through September 1998 **Interventions:** arthroscopic debridement versus arthroscopic lavage versus placebo surgery **Sample size calculations:** a total sample of 180 participants (60 participants in each arm) would provide 90% statistical power and a two-sided type I error of 0.04 to detect a moderate effect size (0.55) between the placebo group and the combined arthroscopic-treatment groups **Analysis:** no information on intention-to-treat analysis or analysis on data from those who did not complete the 24-month follow-up
Participants	**Number of participants** • Number of consecutive participants who were eligible: 324 • Number of participants who declined to participate: 144 • Number of participants at enrolment: 180 • Number randomised: 180: 59 in the arthroscopic debridement group, 61 in the lavage group and 60 in the placebo group • Number completing trial: 165: 53 in the arthroscopic debridement group, 56 in the lavage group and 55 in the placebo group • Number included in analysis of primary outcome (KSPS at 24 months): 163: 53 in the debridement group, 55 in the lavage group and 55 in the placebo group **Criteria for defining knee osteoarthritis:** the definition was according to the American College of Rheumatology. The severity of osteoarthritis in the study knee (that with the greatest pain-induced limitation of function) was assessed radiographically and graded on a scale of zero to four. The scores for the three compartments were added together to generate a severity grade of 0 to 12. **Inclusion criteria** • 75 years of age or younger • Had knee OA as defined by the American College of Rheumatology • Reported moderate knee pain on average (> 4 on a visual-analogue scale ranging from 0 to 10) despite maximal medical treatment for at least six months • Not undergone arthroscopy of the knee during the previous two years **Exclusion criteria** • Severity of osteoarthritis at grade 9 or higher • Severe deformity • Serious medical problems **Baseline characteristics** *Arthroscopic debridement*

Moseley 2002 *(Continued)*

- Mean (SD) age: 53.6 (12.2) years
- Proportion (%) of men and women: M/F = 96.6/3.4
- Severity of osteoarthritis in knee (%): mild (30.5), moderate (45.8), severe (23.7)
- Analgesic use (%): non-prescription (64.4), prescription (15.3)
- Mean score on Knee Society Clinical Rating Scale: knee symptoms (51.4), function (57.6)
- Mean (SD) AIMS score: pain 59.3 (22.2)
- Mean (SD) Psychological attributes: anxiety 28.4 (22.4), depression 22.0 (35.3), expectations for benefit 3.6 (1.1), optimism 73.7 (17.1), satisfaction with general health 46.5 (24.8), social functioning 67.6 (25.2), somatisation 10.0 (10.7), stress 27.9 (18.8), vitality 57.7 (19.3)

Arthroscopic Lavage

- Mean (SD) age: 51.2 (10.5) years
- Proportion (%) of men and women: M/F = 88.5/11.5
- Severity of osteoarthritis in knee (%): mild (27.9), moderate (45.9), severe (26.2)
- Analgesic use (%): non-prescription (67.2), prescription (21.3)
- Mean score on Knee Society Clinical Rating Scale: knee symptoms (50.2), function (62.4)
- Mean (SD) AIMS score: pain 59.3 (16.7)
- Mean (SD) Psychological attributes: anxiety 30.2 (19.9), depression 28.1 (37.2), expectations for benefit 3.5 (0.9), optimism 74.5 (19.4), satisfaction with general health 43.7 (22.4), social functioning 60.3 (23.9), somatisation 9.6 (12.4), stress 26.1 (18.2), vitality 52.7 (19.7)

Placebo Surgery

- Mean (SD) age: 52 (11.1) years
- Proportion (%) of men and women: M/F = 93.3/6.7
- Severity of osteoarthritis in knee (%): mild (28.3), moderate (46.7), severe (25.0)
- Analgesic use (%): non-prescription (70.0), prescription (21.7)
- Mean score on Knee Society Clinical Rating Scale: knee symptoms (49.4), function (62.2)
- Mean (SD) AIMS score: Pain 59.5 (18.5)
- Mean (SD) Psychological attributes: anxiety 27.0 (21.0), depression 20.0 (32.0), expectations for benefit 3.5 (1.0), optimism 72.6 (21.0), satisfaction with general health 39.3 (25.1), social functioning 65.5 (25.6), somatisation 11.3 (12.7), stress 28.4 (19.7), vitality 54.8 (21.0)

Pre-treatment group differences: although baseline characteristics appeared to be similar across all three groups, the lavage group showed higher scores for depression compared to the other two groups and the use of prescription analgesics was lower in the debridement group.

Interventions

One orthopedic surgeon (board-certified, fellowship-trained in arthroscopy and sports medicine, in practice for 10 years in academic medical centre) performed all the operations. Post-operative care was delivered according to a protocol specifying that all participants should receive the same walking aids, graduated exercise program and analgesics. The protocol was not provided in the publication or supplementary appendix.

Arthroscopic debridement
After diagnostic arthroscopy, the joint was lavaged with at least 10 litres of fluid, rough articular cartilage was shaved (chrondroplasty was performed), loose debris was removed, all torn or degenerated meniscal fragments were trimmed, and the remaining meniscus was smoothed to a firm and stable rim. No abrasion arthroplasty or microfracture was performed. Typically, bone spurs were not removed, but any spurs from the tibial spine area that blocked full extension were shaved smooth. Participants received standard general anaesthesia with endotracheal intubation.

Arthroscopic lavage

After diagnostic arthroscopy, the joint was lavaged with at least 10 litres of fluid. Anything that could be flushed out through arthroscopic cannulas was removed. Normally, no instruments were used to mechanically debride or remove tissue. However, if a mechanically important, unstable tear in the meniscus (e.g. a displaced “bucket-handle” tear) was encountered, the torn portion was removed and the re-

Moseley 2002 *(Continued)*

maining meniscus was smoothed to a firm, stable rim. No other debridement was performed. Participants received standard general anaesthesia with endotracheal intubation.

Placebo surgery

Simulated debridement with three 1-cm skin incisions but without insertion of the arthroscope. The surgeon asked for all instruments and manipulated the knee as if arthroscopy were being performed. Saline was splashed to simulate the sounds of lavage. No instrument entered the portals for arthroscopy. The participant was kept in the operating room for the amount of time required for a debridement. Participants received a short-acting intravenous tranquilliser and an opioid and spontaneously breathed oxygen-enriched air. Participants spent the night after the procedure in the hospital and were cared for by nurses who were unaware of the treatment-group assignment.

Outcomes

Outcomes were measured at baseline and at 2 and 6 weeks, and at 3, 6, 12, 18 and 24 months after the procedure.

Primary outcome

Pain in the study knee assessed by a 12-item self-reported Knee-Specific Pain Scale (KSPS) at 24 months' follow-up. Scores range from 0 to 100 with higher scores indicating more severe pain.

Secondary outcomes (measured at all time points)

- A 4-item pain subscale of the Arthritis Impact Measurement Scales (AIMS2-P) with higher scores indicating more severe pain. Scores were transformed into scores on a scale from 0 to 100.
- A 2-item pain subscale of the Medical Outcomes Study 36-item Short-Form General Health Survey (SF-36-P), where higher scores indicate less severe pain. Scores were transformed into scores on a scale from 0 to 100.
- A 5-item walking–bending subscale from the AIMS2 (AIMS2-WB), transformed into scores on a scale from 0 to 100, with higher scores indicating more limited physical function.
- A 10-item physical-function subscale from the SF-36 (SF-36-PF), transformed into scores on a scale from 0 to 100, with higher scores indicating better function.
- The Physical Functioning Scale (PFS) which records the amount of time in seconds required to walk 30 metres and to climb up and down a flight of stairs as quickly as possible. Longer time (in seconds) indicates poorer functioning.

Outcomes included in this review at 3 and 6 months and 2 years

- Pain - SF-36 pain subscale (0 to 100, lower score = less pain)
- Function - SF-36 physical function subscale (0 to 100, higher score = better function)

NB. Adverse events could not be included as the details of the group to which the adverse events belonged were not reported.

Quality of life outcome data could not be included as SF-36 Mental Component Summary (MCS) scores were not reported.

Notes

Funding: supported by a grant from the Department of Veterans Affairs

Trial registration: not reported

Serious adverse events: none reported

Other adverse events

- Incisional erythema - 1 participant
- Calf swelling in the leg that had undergone surgery; venography was negative for thrombosis - 1 participant

Details of the group to which these participants belonged were not reported.

Knee surgery (replacement or osteotomy): not reported

Moseley 2002 *(Continued)*

Progression of knee OA: not reported

Withdrawals: 6/59 in the arthroscopic debridement group, 5/61 in the lavage group and 5/60 in the placebo group

Data analysis: we included arthroscopic debridement versus placebo in Analysis 1.1 and Analysis 1.2 and we excluded the lavage treatment arm from this review. We included the SF-36-P and SF-36-PF scores for pain and function, respectively, but other measures for these outcome domains were available.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants were randomly assigned using a stratified randomisation process in blocks of six based on the severity of osteoarthritis (grade 1, 2 or 3; grade 4, 5 or 6; and grade 7 or 8) using sealed, sequentially-numbered, stratum-specific envelopes. Probably low risk
Allocation concealment (selection bias)	Low risk	Sealed, sequentially-numbered, stratum-specific envelopes containing treatment assignments were prepared by research staff and handed to the surgeon in the operating suite.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding of participants was done ("the treatment assignment was not revealed to the patient"). Study personnel who were unaware of the treatment-group assignments performed all post-operative outcome assessments; the operating surgeon did not participate in outcome assessment. To assess whether participants remained unaware of their treatment-group assignment, they were asked at each follow-up visit to guess which procedure they had undergone. Participants in the placebo group were no more likely than participants in the other two groups to guess that they had undergone a placebo procedure. For example, at 2 weeks, 13.8% of participants in the placebo group guessed that they had undergone a placebo procedure, and 13.2% of participants in the lavage and debridement groups guessed that they had undergone a placebo procedure.
Blinding of outcome assessor Self-reported outcomes	Low risk	Blinding of participants was done and there is low risk of bias in the measurement of pain and function.
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	Knee replacement not measured.
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	6/59 (10%) in the debridement group, 5/61 (8%) in the lavage group and 5/60 (8%) in the placebo group did not complete the study. No data were given on the reasons for loss to follow-up or withdrawals.
Selective reporting (reporting bias)	Low risk	Trial registration not done and protocol not available. The published article included results on all study outcomes as described in methods
Other bias	Low risk	No other bias apparent

Osteras 2012

Study characteristics

Osteras 2012 *(Continued)*

Methods	**Study design:** multicentre, two-arm, randomised controlled trial **Setting:** two orthopedic clinics in two hospitals in Norway **Trial time period:** participants were recruited over a period of one year; however, the exact time frame was not mentioned. **Interventions:** arthroscopic surgery versus exercise **Sample size calculations:** based on a pre-determined difference between treatment groups of 20% change in pain on a 10-cm visual analogue scale and a standard deviation of 1.5 cm, 10 participants were required in each group to have 80% power to detect the 20% difference as statistically significant at the level of $P < 0.05$. **Analysis:** intention-to-treat
Participants	**Number of participants** • Number of participants screened: 29 • Number of participants at enrolment: 17 • Number randomised: 17: 8 in the arthroscopic meniscectomy group and 9 in the medical exercise therapy group • Number included in analyses: 17: 8 in the arthroscopic meniscectomy group and 9 in the medical exercise therapy group (at 3 months) **Inclusion criteria** • Aged 35 to 60 years • Knee pain for more than 3 months • Eligible for an arthroscopic partial meniscectomy • MRI showing a degenerative meniscus tear **Exclusion criteria** • Anterior cruciate ligament (ACL) rupture for individuals requiring acute trauma surgeries • High-energy traumas with ligament injuries • Osteoarthritis grades 3 to 4 (Kellgren-Lawrence classification) • Haemarthroses and acute cases of locking knee • Symptomatic pain in contrary extremities • Other musculoskeletal comorbidities severely affecting lower extremity muscle function that override the symptoms from the knee • Comorbidities excluding physical activities and exercise • Not able to speak or read the language of interest **Baseline characteristics** *Arthroscopic partial meniscectomy* • Mean (SD) age: 52.7 (7.2) years • Mean (SD) body weight: 82.4 (10.9) • Mean (SD) duration of symptoms: 2.1 (1.7) • Mean (SD) stage of arthritis: 0.9 (1.0) • No. (%) of men: 5 (62.5) • Mean (SD) VAS: 3.7 (0.9) • Mean (SD) 5RM (Repetition Max): 8.6 (5.4) • Mean (SD) KOOS: 48.4 (25.6) • Mean (SD) Hospital Anxiety and Depression Scale (HADS) Anxiety: 4.0 (2.6) • Mean (SD) HADS Depression: 5.0 (2.5)

Osteras 2012 *(Continued)*

Medical exercise therapy

- Mean (SD) age: 47.0 (10.4) years
- Mean (SD) body weight: 79.8 (7.5)
- Mean (SD) duration of symptoms: 1.6 (1.2)
- Mean (SD) stage of arthritis: 0.6 (0.7)
- No. (%) of men: 8 (88.9)
- Mean (SD) VAS: 3.5 (1.7)
- Mean (SD) 5RM: 12.4 (6.1)
- Mean (SD) KOOS: 51.4 (24.4)
- Mean (SD) HADS Anxiety: 4.5 (3.1)
- Mean (SD) HADS Depression: 5.0 (2.9)

Pre-treatment group differences: there were no obvious differences in baseline characteristics between the two groups.

Interventions	**Arthroscopic surgery** Arthroscopic partial meniscectomy. A standard arthroscopic partial meniscectomy was applied as a surgical intervention, which was carried out at two hospitals in Trondheim, Norway, and performed on participants who fulfilled inclusion criteria and were randomised to surgical treatment. Normal procedures for this surgery at the respective hospitals were followed, the protocols did not differ between the hospitals, and there were two surgeons involved. **Exercise** Supervised (medical) exercise therapy. The exercise program was developed for this particular study, with a focus on co-ordination and muscle function training, along with pain modification exercise therapy. The program was for 3 months, and participants exercised 3 times per week. Each treatment in the exercise group started with 15 to 20 minutes of aerobic work on a stationary ergometer cycle. After 4 exercises each of 3 sets of 30 repetitions halfway through the exercise program, the subjects cycled for 10 minutes and again after the last 4 exercises, the participants did another 10 minutes on a stationary ergometer cycle.
Outcomes	Outcomes were measured at baseline and at 3 months of follow-up. **Primary outcome** Pain in the last week measured with a visual analogue scale (VAS) at rest and recorded on a 0 to 10 cm line (0 = no pain and 10 = maximal pain) **Secondary outcomes** • Knee injury and Osteoarthritis Outcome Score (KOOS) comprising 5 subscales - pain, symptoms, ADL, sports/rec, QoL; scores range from 0 to 100 where 100 indicates good knee function • Hospital Anxiety and Depression Scale (HADS), a self-screening questionnaire for depression and anxiety scores ranged from 0 to 21 for anxiety and 0 to 21 for depression; a lower score indicates a better clinical status. • 5RM - dynamic quadriceps muscle strength was measured with a leg extension bench and a protocol in which the participants lifted a weight with a maximum external load using 5 repetitions. **Outcomes used in this review at 3 months** • Pain measured on VAS at rest (0 to 10, lower score = less pain) • Function measured on KOOS total score (0 to 100, higher score = better function) (KOOS ADL subscale was not reported separately)
Notes	**Funding**: no information on funding was provided in this article **Trial registration:** not reported

Osteras 2012 *(Continued)*

Adverse events: unclear if measured; not reported

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: not reported

Withdrawals: unclear; pilot study results published in conference abstract (Osteras 2011, secondary publication of Osteras 2012) includes more participants

Conference abstract (n = 22 participants); not reported if same participants as in Osteras 2012

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information on sequence generation process provided
Allocation concealment (selection bias)	Unclear risk	"The randomisation procedure was concealed from the experimenters and treating physiotherapist". There was no information on how the allocation was concealed.
Blinding of participants and personnel (performance bias) All outcomes	High risk	The interventionists were blinded to the group assignment; however, participants were unable to be blinded due to the nature of the intervention (surgery versus physiotherapy).
Blinding of outcome assessor Self-reported outcomes	High risk	Participants were not blinded and there is risk of bias in the measurement of pain, other symptoms, activities of daily living, functioning in sport and recreation, knee-related quality of life, depression and anxiety.
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	High risk	As the outcome assessor was not blinded to the intervention group assignment, there is risk of bias in the measurement of the quadriceps muscle strength tests.
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Different numbers of participants were included in two reports (n = 17 versus n = 22). These are likely to be from a single study (trial registration not reported).
Selective reporting (reporting bias)	Unclear risk	Trial registration not done. While outcomes described in methods were reported in the results of the main publication, there is a discrepancy in participant numbers across two reports so the risk of bias is unclear.
Other bias	Unclear risk	Unclear if there was an unplanned interim analysis performed

Roos 2018

Study characteristics

Methods

Study design: multicentre, prospective, double-blind, randomised, placebo-controlled trial

Setting: outpatient departments of the orthopaedic clinics in Region Zealand; Slagelse Hospital; Næstved Hospital, Denmark

Trial time period: 21 February 2011 to March 2015

Interventions: arthroscopic surgery versus placebo surgery

Roos 2018 *(Continued)*

Sample size calculations: 36 individuals per group would be required to obtain a power of at least 80% to detect a minimal important change (MIC) of 10 KOOS 5 score units, assuming a common standard deviation of 15. The study team decided to include 80 individuals in total (40 participants in each group), allowing for a 10% dropout rate.

Analysis: intention-to-treat analysis was planned and both a per-protocol and ITT analysis (using best observation carried forward for missing data) were executed.

Participants

Number of participants

- Number screened: 586 (351 not eligible, 135 declined to participate)
- Number of participants at enrolment: 100 (32 showed no meniscal lesion in MRI, 24 declined participation)
- Number randomised: 44: 22 in arthroscopic surgery group and 22 in the placebo surgery group, 8 participants crossed over from the placebo surgery group to the arthroscopic surgery group
- Number included in analyses: 42 (21 in each group at 3 months); 22 in the arthroscopic surgery group and 20 in the placebo surgery group at 24 months

Inclusion criteria

- Participants with knee pain for more than 2 months without significant trauma (only the index knee included)
- MRI-confirmed medial meniscus lesion (participants with a medial meniscus lesion and a lateral lesion also included)
- Age 35 to 55 years
- Eligible for outpatient surgery

Exclusion criteria

- Participants in need of acute surgery e.g. locking knees, high-energy trauma
- Symptoms associated with other musculoskeletal comorbidities overriding the symptoms of the knee
- Grade 3 or 4 knee OA on the Kellgren-Lawrence classification
- Knee surgery within the last 2 years
- Obese participants with BMI > 35
- Participants with ischaemic heart disease
- Participants with diabetic late-complications
- Participants with thrombophilia
- Pregnancy
- Participants unable to speak Danish
- Participants with drug or alcohol abuse

Baseline characteristics

Arthroscopic partial meniscectomy (n = 22)

Mean (SD) age: 47.2 (5.9)

No. (%) male/female: 13/9 (59/41)

Mean (SD) BMI: 27.6(3.6)

No. (%) Kellgren-Lawrence grading: 0 - 9 (41), 1 - 9 (41), 2 - 4 (18), 3 - 0, 4 - 0

No. (%) joint line tenderness: 21 (100)

No. (%) positive McMurray test: 17 (81)

No. (%) swelling present: 11 (52)

No. (%) full knee extension: 18 (86)

Roos 2018 *(Continued)*

No. (%) small extension deficit < 10º: 3 (14)

No. (%) full knee flexion: 17 (81)

No. (%) small flexion deficit < 10º: 3 (14)

No. (%) large flexion deficit: 1 (5)

Median (IQR) duration of pain (months): 5 (2-6)

Mean (SD) $KOOS_5$ (0–100 worst to best): 51.2 (15.6)

Mean (SD) KOOS Pain (0–100): 55.1 (15.4)

Mean (SD) KOOS Symptoms (0–100): 62.8 (17.7)

Mean (SD) KOOS Function in daily living (0–100): 64.9 (19.9)

Mean (SD) KOOS Function in sport and recreation (0–100): 35.0 (23.0)

Mean (SD) KOOS Knee-related quality of life (0–100): 38.7 (15.4)

Mean (SD) EQ-5D VAS score: 69 (14)

Mean (SD) EQ-5D 3L index value (0–1): 0.749 (0.108)

Mean (SD) SF-36 Physical Component Summary (0–100): 38 (10)

Mean (SD) SF-36 Mental Component Summary (0–100): 59 (7)

Placebo surgery (n = 22)

Mean (SD) age: 46.4 (5.5)

No. (%) male/female: 10/12 (45/55)

Mean (SD) BMI: 26 (3.9)

No. (%) Kellgren-Lawrence grading: 0 - 10 (45), 1 - 8 (36), 2 - 4 (18), 3 - 0, 4 - 0

No. (%) joint line tenderness: 20 (91)

No. (%) positive McMurray test: 17 (77)

No. (%) swelling present: 8 (36)

No. (%) full knee extension: 21 (95)

No. (%) small extension deficit < 10º: 1 (5)

No. (%) full knee flexion: 14 (64)

No. (%) small flexion deficit < 10º: 6(27)

No. (%) large flexion deficit: 2 (9)

Median (IQR) duration of pain (months): 3.5 (2.0 – 6.0)

Mean (SD) $KOOS_5$ (0–100): 44.8 (19.9)

Mean (SD) KOOS Pain (0–100): 45.9 (22.0)

Mean (SD) KOOS Symptoms (0–100): 59.9 (20.6)

Mean (SD) KOOS Function in daily living (0–100): 56.5 (22.3)

Mean (SD) KOOS Function in sport and recreation (0–100): 25.2 (26.3)

Roos 2018 *(Continued)*

	Mean (SD) KOOS Knee-related quality of life (0–100): 36.6 (20.2) Mean (SD) EQ-5D VAS score: 63 (SD not reported) Mean (SD) EQ-5D 3L index value (0–1): 0.642 (SD not reported) Mean (SD) SF-36 Physical component summary (0–100): 35 (SD not reported) Mean (SD) SF-36 Mental component summary (0–100): 57 (SD not reported) **Pre-treatment group differences:** the KOOS scores ($KOOS_5$ and subscales) were higher in the arthroscopic group compared to the placebo surgery group.
Interventions	**Arthroscopic surgery** The arthroscopic partial meniscectomy was performed on an outpatient basis under general anaesthesia combined with local anaesthesia. Arthroscopic surgery was performed by experienced surgeons in their final year of residency or attending orthopaedic surgeons. Two standard portals on the lateral and medial sides of the ligamentum patella were created but no outflow cannula inserted. An arthroscope was used with a pressure-controlled irrigation system. Tourniquet use was at the discretion of the surgeon. The strategy for the meniscectomy was to preserve as much tissue as possible. A standard operation protocol was used to document possible findings in cartilage, ligaments, synovium and the medial and lateral menisci. The type, and extent of meniscus lesion was registered and changes in the articular cartilage was classified according to the ICRS (International Cartilage Repair Society) classification. **Placebo surgery** The placebo procedure (skin incision only) was performed under the same conditions as the arthroscopic surgery. Participants were fully sedated with general anaesthesia. Local anaesthetic was applied and two skin incisions made at the same locations and of the same size as in the arthroscopic surgery group. Then the knee was manipulated as if a real arthroscopy was performed, the spillage of water and all other equipment needed for an arthroscopy was used. No instruments entered the arthroscopy portals to avoid the possibility of deep infection, osteochondral lesions or unwanted interventions by the surgeon. A pre-recorded video of a standard arthroscopic partial meniscectomy was planned to be played during the placebo procedure (but was not done). All participants in both intervention groups were given a folder including an exercise program for postoperative participants after knee arthroscopy or placebo procedure. The exercise program included 7 different non-weight-bearing exercises to improve lower extremity function and knee range of motion (for the first week after surgery) and a further 3 weight-bearing exercises thereafter. The exercises were for the participants to carry out at home and were recommended to be performed 10 to 15 times three times daily. The participants were also advised to start unloaded cycling, swimming or walking after 1 week, and jogging or loaded cycling after 2 to 3 weeks.
Outcomes	Outcomes were measured at baseline, 3 and 24 months. **Primary outcome** • $KOOS_5$ composite score at 24 months. $KOOS_5$ is derived from the Knee injury and Osteoarthritis Outcome Score (KOOS) and is calculated as a mean of the 5 subscale scores: (KOOS pain + KOOS symptoms + KOOS ADL + KOOS sport and rec + KOOS QoL / 5). Scores for each subscale range from 0 (extreme symptoms) to 100 (no symptoms) **Secondary outcomes** • KOOS individual subscales - pain, symptoms, ADL, sports/rec, QoL. Scores for each subscale range from 0 (extreme symptoms) to 100 (no symptoms) • Global Perceived Effect score - a rating of overall improvement in knee symptoms after the operation (7-point scale ranging from 'much worse' to 'much better'). A clinically important change is considered an improvement or worsening of at least 2 steps on the scale • SF-36 Physical Component Summary (PCS) and Mental Component Summary (MCS) (0 to 100, higher = better)

Roos 2018 *(Continued)*

- EQ-5D generic quality of life using 2 scales: EQ-5D VAS global assessment of disease status scale (0 to 100, higher = better) and EQ-5D 3L descriptive system (0-1, higher = better; contains 5 domains: mobility, self-care, usual activities, pain/discomfort, anxiety/depression; each rated on 3-point scale: no problems, some problems, severe problems)
- Single leg hop test (best of 3 trials, measured in centimetres)
- Knee-bend test (maximum number achieved in 30 seconds)
- Isometric knee extension strength (highest of 3 contractions) measured sitting using a dynamometer
- Radiography of knees to assess possible onset or progress of knee OA (score assigned based on joint space width and presence of osteophytes using standard atlas), at baseline and 5 years
- Self-efficacy using the modified Danish Arthritis Self-Efficacy Scale
- Participant expectations (single item)
- Adverse events (e.g. superficial infection, nerve or vessel injury, deep infection, compartment syndrome, DVT, MI, stroke and death)

Outcomes used in this review at 3 and 24 months

- Pain - KOOS pain subscale (0 to 100, higher score = less pain) (data supplied by trial authors) (we multiplied the mean values by –1 so that lower score = less pain, as per *Cochrane Handbook* guidance)
- Function - KOOS ADL subscale (0 to 100, higher score = better function)
- Knee-specific health-related quality of life - KOOS QoL subscale (0 to 100, higher score = better QoL)
- Generic health-related quality of life - SF-36 MCS (0 to 100, higher = better QoL)
- Participant reported treatment success - Global perceived effect (rating of 'better' or 'much better') at 24 months (data supplied by trial authors)
- Total adverse events at 24 months

Serious adverse events could not be included as the details of the group to which some serious adverse events belonged were not reported.

Notes

Funding: University of Southern Denmark, Odense; the Orthopedic Departments of Slagelse and Næstved Hospital; the Research Unit of Hospital South, Region Zealand; Edith and Henrik Henriksens Memorial Fund; Region Zealand Health Scientific Research Fund; Research Fund of Hospital South

Clinical trial registration: NCT01264991

Adverse events: 4/22 knee-related adverse events in the surgery group, of which two were serious (two re-arthroscopies comprising one partial meniscectomy and one anterior cruciate ligament reconstruction), two were not serious (one cutaneous nerve lesion and one mild knee swelling). No reported knee-related adverse events reported in the placebo group.

Total other adverse events - 7 adverse events in 5 participants (2 participants in the surgery group and 3 participants from placebo group). The nature of events included: chest pain, finger injury, nausea, dizziness and kidney stone, and included two regarded as serious (abdominal surgery and malignant melanoma). Details of the group to which some serious adverse events belonged were not reported.

Knee surgery (replacement or osteotomy): none

Progression of knee OA: not reported

Withdrawals: 1/22 in the arthroscopy group and 1/22 in the placebo group at 3 months, and 2/22 participants in the placebo group only at 24 months

Post-protocol changes: a pre-recorded video of a standard arthroscopic partial meniscectomy was planned to be played during the procedure for both the surgery and placebo group, but was not performed in the placebo group.

Statistical power: trial was underpowered due to inability to recruit 40 participants per group as planned

Roos 2018 *(Continued)*

Data analysis: SMD for generic quality of life at > 6 months up to 2 years (Analysis 1.4) back-translated to SF-36 PCS by multiplying the SMD by the standard deviation at baseline in the surgery group (as standard deviation in the placebo group was not reported) (SD = 10)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Used a computer-generated table of random numbers, prepared by an external co-investigator
Allocation concealment (selection bias)	Low risk	Used consecutively-numbered, sealed envelopes stored in a briefcase outside the operating theatre
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants were blinded to group allocation at short-term follow-up, there was low risk of bias at primary time point. Blinding broken for 16 participants (36%; 6/22 from arthroscopy group and 10/22 from placebo group) prior to 2-year follow-up. Prior to 3 months, two additional participants from the arthroscopy group (due to adverse events) and one from the placebo group (persisting pain) were unblinded by the treating surgeon. Between 3 and 24 months, two from the arthroscopy group and eight from the placebo group were unblinded by the treating surgeon because of persisting pain. In total, 10/22 in the arthroscopy group and 19/22 in the placebo group were unblinded. Operating surgeons and theatre staff were not blinded.
Blinding of outcome assessor Self-reported outcomes	Low risk	Low risk at short-term follow-up, unclear if knowledge of treatment in 6/22 and 10/22 influenced assessment of pain and function at 2 years
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	Outcome assessors were blinded to group allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Self-reported outcome data were missing for 2/22 participants in the placebo group at baseline. Data missing for 1/22 in the arthroscopy group and 1/22 in the placebo group at 3 months. 2/22 participants in the placebo group were lost to follow-up at 2 years (crossed over to arthroscopic surgery between 3 and 24 months)
Selective reporting (reporting bias)	Low risk	3-month follow-up data were provided upon request from the authors
Other bias	Low risk	No other bias apparent

Saeed 2015

Study characteristics

Methods	**Study design:** single centre, two-arm, randomised controlled trial **Setting:** Department of Orthopedics, Ch. Rehmat Ali Memorial Trust Hospital attached to Continental Medical College Lahore, Pakistan **Trial time period:** January 2012 to December 2014

Saeed 2015 *(Continued)*

Interventions: arthroscopic surgery versus five intra-articular hyaluronic acid injections given at weekly intervals

Sample size calculations: not reported

Analysis: intention-to-treat

Participants	**Number of participants** • Number of participants screened: not reported • Number of participants at enrolment: 120 • Number randomised: 120: 60 in the arthroscopic debridement group and 60 in the intra-articular hyaluronic acid injection group • Number included in analyses: 120: 60 in the arthroscopic debridement group and 60 in the intra-articular hyaluronic acid injection group **Inclusion criteria** • Age > 40 years • History of pain in the knee joint • Grade 2 & 3 Kellgren-Lawrence grading system **Exclusion criteria** • Age < 40 years • History of injury or accident • Prior intervention with intra-articular glucocorticoid injections within the past three months **Baseline characteristics** *Arthroscopic surgery* • No. (%) women: 48 (80) • No. (%) men: 12 (20) • No. (%) of Kellgren-Lawrence Grade 2 - 38 (63.33), Grade 3 - 22 (36.66) • No. (%) Pain score measured on the Knee Society Score System (0 = severe, 50 = none): 10 = 4 (6.7), 20 = 42 (70), 30 = 14 (23.3) *Intra-articular hyaluronic acid injections* • No. (%) women: 50 (83.3) • No. (%) men: 10 (16.6) • No. (%) of Kellgren-Lawrence Grade 2 - 36 (60), Grade 3 - 24 (40) • No. (%) Pain score measured on the Knee Society Score System (0 = severe, 50 = none): 10 = 8 (13.4), 20 = 26 (43.3), 30 = 26 (43.3) **Pre-treatment group differences**: there were some differences between groups in pain at baseline: in the intra-articular hyaluronic acid injection group, 13% of participants had pain score 10, 43% had 20 and 43% had 30, while in the arthroscopic group, 7% of participants had pain score 10, 70% had 20 and 23% had 30.
Interventions	**Arthroscopic surgery** Arthroscopic debridement. Participants were admitted to the hospital, arthroscopic debridement was performed in the operation theatre by using two portals in all cases under spinal anaesthesia. Participants were discharged on the next day. Monitoring of electrocardiogram (ECG) and blood pressure was standard in all cases during the entire duration of the procedure. All the debridements were done by a single surgeon to minimise the bias for the study. **Intra-articular hyaluronic acid injections**

Saeed 2015 *(Continued)*

	Participants were injected with intra-articular hyaluronic acid after being given intradermal anaesthesia. The injections were given weekly for five weeks with a 24-gauge needle under strict aseptic conditions in the operation theatre as an outpatient. In case of joint effusion, aspiration was done before the injection to prevent dilution of the injection.
Outcomes	Outcome were assessed at baseline and at 1, 3 and 6 months' follow-up **Outcomes** • Knee Society Score System (KSSS) assessing knee pain scores; range from 0 - severe to 50 - none. This study categorised the scores into KSSS 20, KSSS 30, KSSS 40, KSSS 45 • Adverse events **Outcomes used in this review at 3 and 6 months** • Pain measured on KSSS, score 30 or higher • Total adverse events
Notes	**Funding:** not reported **Trial registration:** not reported **Adverse events:** *Arthroscopic debridement group* **Serious adverse events:** none reported **Other adverse events:** *No.(%)*: 13 (26) *Nature of event:* pain and mild effusion **Total adverse events:** No.(%): 13 (26) *Intra-articular hyaluronic acid injection group* **Serious adverse events:** none reported **Other adverse events:** *No.(%)*: 8 (13.4) *Nature of event:* pain at injection site **Total adverse events:** No.(%): 8 (13.4) The frequency of adverse events in the arthroscopy group was higher than the injection group **Knee surgery (replacement or osteotomy):** not reported **Progression of knee OA:** not reported **Withdrawals:** none

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Described as 'randomized experimental study' but no information about sequence generation process provided

Saeed 2015 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	No information on whether allocation concealment was done or not
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding of participants and study personnel was not reported. Probably not done due to the nature of the intervention
Blinding of outcome assessor Self-reported outcomes	High risk	Risk of detection bias in assessment of participant-reported knee pain and adverse events as blinding of participants was not reported but probably not done
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	Blinding of study personnel not reported but adverse effects unlikely to be influenced by lack of blinding
Incomplete outcome data (attrition bias) All outcomes	Low risk	No withdrawals or loss to follow-up
Selective reporting (reporting bias)	High risk	Trial registration not done and protocol not available. The authors measured overall knee function (using the Knee Society Score System), but it is unclear if they reported overall scores or a pain subscore.
Other bias	Low risk	No other bias apparent

Sihvonen 2013

Study characteristics	
Methods	**Study design:** multicentre, randomised, parallel-arm, double-blind, placebo-controlled trial **Setting:** five orthopedic clinics in Finland **Trial time period:** participants were enrolled between December 2007 and January 2013 **Interventions:** arthroscopic partial meniscectomy (APM) plus home exercises versus placebo surgery plus home exercises **Sample size calculations:** a total sample of 134 participants with 40, 54 and 40 participants per group for the Lysholm score, WOMET score and pain assessment, respectively, with 80% power to show a clinically meaningful advantage of APM over placebo, based on a two-sided type 1 error rate of 5%. Anticipating a loss to follow-up of at least 20%, the study planned to recruit 70 participants per group. **Analysis:** intention-to-treat
Participants	**Number of participants** • Number of participants screened: 205 (45 were excluded before arthroscopy and 14 excluded after diagnostic arthroscopy) • Number of participants at enrolment: 146 • Number randomised: 146 participants: 70 to the arthroscopic partial meniscectomy group and 76 to the placebo surgery group • Number included in analyses: 146: 70 in the arthroscopic partial meniscectomy group and 76 in the placebo surgery group (at 12 months - primary time point)

Sihvonen 2013 *(Continued)*

Criteria for defining study participants: those who have knee symptoms consistent with a degenerative medial meniscus tear and no knee osteoarthritis

Inclusion criteria

- 35 to 65 years of age
- Persistent (> 3 months) pain on the medial joint line of the knee
- Pain that can be provoked by palpation or compression of the joint line or a positive McMurray sign
- Tear of the medial meniscus on MRI
- Degenerative injury to the medial meniscus confirmed at arthroscopy

Exclusion criteria

- Acute, trauma-induced onset of symptoms.
- Locking or painful snapping of the knee joint
- A surgical operation performed on the affected knee
- Osteoarthritis of the medial compartment of the knee (determined by clinical criteria of the Amercican College of Rheumatology)
- Osteoarthritis on knee radiographs (Kellgren-Lawrence grade > 1)
- Acute (within the previous year) fractures of the knee
- Decreased range of motion of the knee
- Instability of the knee
- MRI assessment showing a tumour or any other complaint requiring surgical or other means of treatment
- Arthroscopic assessment showing anything other than a degenerative tear of the medial meniscus requiring surgical intervention

Baseline characteristics

Arthroscopic partial meniscectomy

- Mean (SD) age: 52 (7) years
- Number of men and women: M/F = 42/28
- Mean (SD) body-mass index: 26.9 (4)
- Median (range) duration of medial knee pain: 10 (3-50) months
- Onset of symptoms - no. (%) : gradual 48 (69) after exercise 12 (17) sudden 10 (14)
- Positive result of McMurray test — no. (%) 16 (23)
- Kellgren–Lawrence grade — no. (%): 0- 35 (50); 1 - 35 (50)
- Symptoms of catching or locking no.(%): 32(46)
- Pain provoked by forced flexion, causing compression at the medial tibiofemoral joint line — no. (%): 50 (71)
- Pain provoked by palpation at the medial tibiofemoral joint line — no. (%): 63 (90)
- Mean (SD) Lysholm knee score: 60.2 (14.7)
- Mean (SD) WOMET score: 56.4 (17.3)
- Mean (SD) for knee pain (VAS): after exercise 5.8 (2.0) at rest 4.1 (2.3)
- Mean (SD) 15D score: 0.90 (0.06)

Placebo surgery

- Mean (SD) age: 52 (7) years
- Number of men and women: M/F = 47/29
- Mean (SD) body-mass index: 27.9 (4)
- Median (range) duration of medial knee pain: 10 (3-47) month
- Onset of symptoms - no. (%) : gradual 48 (63) after exercise 14 (18) sudden 14 (18)
- Positive result of McMurray test — no. (%) 15 (20)
- Kellgren–Lawrence grade — no. (%): 0- 36 (47); 1 - 40 (53)
- Symptoms of catching or locking no.(%): 37(49)

Sihvonen 2013 *(Continued)*

- Pain provoked by forced flexion, causing compression at the medial tibiofemoral joint line — no. (%): 59 (78)
- Pain provoked by palpation at the medial tibiofemoral joint line — no. (%): 74 (97)
- Mean (SD) Lysholm knee score: 60.1 (14.6)
- Mean (SD) WOMET score: 52.8 (18.1)
- Mean (SD) for knee pain (VAS): after exercise 6.1 (2.0) at rest 4.4 (2.4)
- Mean (SD) 15D score: 0.89 (0.06)

Pre-treatment group differences: there were no differences in the baseline characteristics between the two groups.

Interventions	Arthroscopic examination of the knee was first performed in all participants with the use of standard anterolateral and anteromedial portals and a 4-mm arthroscope. The orthopedic surgeon evaluated the medial, lateral and patellofemoral joint compartments and graded the intra-articular pathologic changes. **Arthroscopic surgery** Arthroscopic partial meniscectomy plus home exercises. The damaged and loose part of the meniscus tissue was removed with arthroscopic instruments (mechanised shaver and meniscal punches) until solid meniscus tissue was reached. The meniscus was then probed to ensure that all loose and weak fragments and unstable meniscus tissue had been successfully resected, preserving as much of the meniscus tissue as possible. **Placebo surgery** Placebo surgery plus home exercises. A standard arthroscopic partial meniscectomy procedure was simulated. The surgeon asked for all instruments and manipulated the knee as if an arthroscopic partial meniscectomy was being performed. The mechanised shaver (without the blade) was pushed firmly against the patella, outside of the knee, to mimic as closely as possible the feelings and sounds of the normal use of the arthroscopic shaver. Further, to simulate the sounds of normal arthroscopic partial meniscectomy, suction was also used to drain the joint and saline was splashed. The participant was kept in the operating room for the amount of time required to perform an actual arthroscopic partial meniscectomy. All procedures were standardised and recorded on video. In both arthroscopic partial meniscectomy and placebo surgery groups, the post-operative care was delivered according to a standard protocol specifying that all participants received the same walking aids and graduated home exercise programme. Participants were instructed to take over-the-counter analgesic agents as required.
Outcomes	Outcomes were assessed at baseline and 2, 6, 12, 24, 36, 48 and 60 months' follow-up. **Primary outcomes** • Knee pain after exercise (during the preceding week) at 12 months (primary time point) after surgery using an 11-point numerical rating scale ranging from 0 (no pain) to 10 (extreme pain) • Lysholm knee score at 12 months (primary time point). Scores range from 0 to 100 with lower scores indicating more severe symptoms • Western Ontario Meniscal Evaluation Tool (WOMET) score at 12 months (primary time point). WOMET is a meniscus-specific health-related quality-of-life instrument with scores ranging from 0 (worst possible situation) to 100 (best possible situation). **Secondary outcomes** • Knee pain after exercise (during the preceding week) at 2, 6, 24 and 60 months after surgery using an 11-point numerical rating scale (0 'no pain' to 10 'extreme pain') • Lysholm knee score at 2, 6, 24 and 60 months after surgery (0 to 100, lower scores = more severe symptoms) • WOMET score at 2, 6, 24 and 60 months after surgery (0 to 100, higher scores = better quality of life) • Knee pain at rest at 12 months, using a numerical rating scale (0 'no pain' to 10 'extreme pain')

Sihvonen 2013 *(Continued)*

- 15D generic quality of life score at 12 months after surgery. Scores range from 1 (full health) to 0 (being dead)
- Serious adverse events at 12, 24 and 60 months, defined as untoward medical occurrences that may or may not have had a causal relationship with the treatment administered. Classified as serious if they necessitated hospitalisation or prolonged inpatient hospital care, or if they were life-threatening or resulted in death
- Treatment success at 12, 24 and 60 months after surgery (measured by participant responses to 3 questions: 'Is your knee better than before the intervention?'; 'Are you satisfied with your knee at present?'; 'Would you choose to be operated on again if you were asked to make the decision now?'). Responses for the first 2 questions were rated on 5-point Likert scales ('much better' to 'much worse'). 'Much better' and 'Better' were considered to indicate improvement, and 'Unchanged', 'Worse', and 'Much worse' were considered to indicate no improvement.)
- Return to previous activities at 24 and 60 months after surgery, rated as yes/no
- Need for subsequent knee surgery (additional arthroscopy, high tibial osteotomy or total knee replacement) at 24 and 60 months after surgery
- Clinical examination at 24 months after surgery including clinical meniscus test, McMurray test, pain provoked by joint line palpation, pain provoked by forced flexion and varus, range of motion, knee crepitus, bony enlargement, effusion, location of pain at palpation and knee stability at 24 months (these outcomes were reported in Sihvonen 2018 (a secondary publication of Sihvonen 2013) but not pre-defined or reported in Sihvonen 2013.
- A cost-utility analysis (based on the participants' general quality of life using the 15D score and the utilisation of healthcare resources) at 12 months after surgery
- Progression of knee osteoarthritis at 5 years - at least one grade progression in radiographic tibiofemoral knee OA on the KL classification (grade 1 (doubtful): doubtful joint space narrowing and possible osteophytic lipping; grade 2 (minimal): definite osteophytes and possible joint space narrowing; grade 3 (moderate): moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends; grade 4 (severe): large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends). Radiographic progression also assessed as being based on the sum of marginal tibiofemoral osteophyte grades and tibiofemoral joint space narrowing grades (according to the atlas developed by OARSI), ranging from 0-18

Outcomes used in this review at 2, 6 and 12 months, and 2 and 5 years

- Pain - pain after exercise on numerical rating scale at 2 and 6 months, and 2 and 5 years (0 to 10, lower score = less pain)
- Function - Lysholm Knee Score at 2 and 6 months, and 2 and 5 years (0 to 100, higher score = better function)
- Generic health-related quality of life - 15D at 12 months (0 to 1, higher score = better)
- Participant reported treatment success at 5 years, measured as number of participants reporting 'much better' or 'better' for item 'Is your knee better than before the intervention?'
- Knee surgery (replacement or osteotomy) at 5 years
- Serious adverse events at 5 years
- Total adverse events at 5 years
- Progression of knee osteoarthritis at 5 years

Notes

Funding: funded by the Sigrid Juselius Foundation, the Competitive Research Fund of Pirkanmaa Hospital District, and the Academy of Finland.

Trial registration: ClinicalTrials.gov number NCT00549172; and NCT01052233 trial number for 10-year follow-up which is ongoing

Adverse events

Arthroscopic partial meniscectomy:

7/70 (10%) participants from the arthroscopic surgery group had a serious knee-related adverse event (3 knee replacement, 4 arthroscopies) and 1/70 (1%) participants from this group had other serious adverse events (1 deep infection of the index knee at 4 months) (8/70 (11.43%) participants)

Sihvonen 2013 *(Continued)*

Placebo surgery:

8/76 (10.52%) participants from the placebo group had serious knee-related adverse events (1 proximal tibial osteotomy, 7 arthroscopic partial meniscectomies) and no other serious adverse events in this group

Knee surgery (replacement or osteotomy) (5-year follow-up)

Arthroscopic partial meniscectomy:

No. (%) of participants = 3/70 (4%) participants from the arthroscopic surgery group had a subsequent knee replacement

Placebo surgery:

1/76 (1%) participants from the placebo surgery group had a subsequent high tibial osteotomy

[trial authors reported 3/68 (4%) in the arthroscopy group and 1/74 (1%) in the placebo group at 5 years. Number receiving allocated intervention was used as denominator in meta-analysis]

Progression of knee OA

At 5 years, 44/74 (59.5%) participants from the placebo group and 48/67 (71.6%) participants from the arthroscopy group had at least one grade progression in radiographic tibiofemoral knee OA on the Kellgren-Lawrence classification.

Withdrawals

At 12 months, there was no loss to follow-up and all randomised participants completed the study. Lysholm Knee Score data for one participant in the placebo surgery group were missing at 6 months' follow-up (no reason given) and data were not imputed. At 24 months, 2 participants from the placebo group were lost to follow-up (one not responding to contact attempts and one deceased). At 60 months, 4 participants (2 placebo and 2 arthroscopic surgery) were lost to follow-up (two not responding and 2 deceased - reasons not given per group)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Stratified randomisation was performed by a statistician using a computer-generated schedule
Allocation concealment (selection bias)	Low risk	Sequentially-numbered, opaque, sealed envelopes were prepared by a statistician with no involvement in the clinical care of participants in the trial
Blinding of participants and personnel (performance bias) All outcomes	Low risk	The operating surgeon and other staff in the operating room were aware of group assignment before the procedure but did not participate in further treatment or follow-up of participants. Participants were blinded. At 12 months, 2/70 (3%) participants in the arthroscopic surgery group and 5/76 (7%) participants in the placebo surgery group reported persistent symptoms after surgery that were sufficiently severe to lead to revealing of the study-group assignment at an average of 8 months after surgery. At 24 months, 5/70 (7%) in the arthroscopic surgery group and 7/74 (9%) participants in the placebo group reported symptoms so severe to lead to unblinding. At 60 months, 8/68 (12%) in the arthroscopic surgery group and 8/74 (11%) participants in the placebo group reported symptoms so severe to lead to unblinding
Blinding of outcome assessor Self-reported outcomes	Low risk	Participants were blinded to the group assignment. Participants in the placebo surgery group were not significantly more likely than participants in arthroscopic surgery group to guess that they had undergone a placebo procedure (47% and 38% respectively, P = 0.39)

Sihvonen 2013 *(Continued)*

Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	Outcome assessors were blinded; low risk of bias for assessment of knee replacement
Incomplete outcome data (attrition bias) All outcomes	Low risk	At 12 months, there was no loss to follow-up and all randomised participants completed the study. Lysholm Knee Score data for one participant in the placebo surgery group were missing at 6 months' follow-up (no reason given) and data were not imputed. However, it is unlikely that this would have had a clinically important impact on the observed effect size. At 24 months, 2 participants from the placebo group were lost to follow-up (one not responding to contact attempts and one deceased). At 60 months, 4 participants (2 placebo and 2 arthroscopic surgery) were lost to follow-up (two not responding and 2 deceased - reasons not given per group)
Selective reporting (reporting bias)	Unclear risk	One primary outcome (WOMET at 12 months) was not pre-specified and was added after data collection but before data analysis (trial registration was amended and revised protocol published). The rationale for adding this primary outcome was provided (the score was validated in the participant population). Two other pre-specified outcomes were changed: (1) pain at rest at 12 months was pre-specified as a primary outcome but moved to a secondary outcome after data collection but before data analysis. The rationale was that reporting both pain at rest and after exercise was 'somewhat ambiguous' and that pain after exercise was the more important of the two; (2) secondary cost-utility analysis based on 15D score and healthcare resource utilisation at 12 months was removed before data analysis (no reason given). The 2-year follow-up paper (Sihvonen 2018) reported the primary time point as 24 months (but protocol states primary time point as 12 months). A new statistical analysis plan for 5- and 10-year follow-up was published in 2020. Results on revised primary and secondary outcomes were reported
Other bias	Low risk	No other bias apparent

Van de Graaf 2018

Study characteristics	
Methods	**Study design:** non-inferiority, parallel-arm, multicentre, randomised controlled trial **Setting:** nine hospitals in the Netherlands **Trial time period:** July 2013 to November 2015 **Interventions:** arthroscopic partial meniscectomy (APM) versus physical therapy (PT) **Sample size calculations:** initially a sample size of 402 participants was estimated to have power of 90%, an α of 0.05 and SD of 20 points; a clinically relevant difference of 8.8 points on the International Knee Documentation Committee (IKDC) 'Subjective Knee Form' was rounded down to a non-inferiority threshold of 8 to increase the power. However, an interim analysis led to recalculation of SD to 18 points resulting in a sample size requirement of 320 participants (120 per group). **Analysis:** intention-to-treat. As-treated analysis was also conducted and results reported in 3 groups - APM group, PT group and delayed APM group i.e. those who were randomised to the PT group but received APM during follow-up.
Participants	**Number of participants** • Number of participants screened: not reported

Van de Graaf 2018 *(Continued)*

- Number of participants at enrolment: 321
- Number randomised: 321 participants: 159 to the arthroscopic partial meniscectomy (APM) group and 162 to the physical therapy (PT) group. 1 participant from each group withdrew immediately after randomisation and their data were not included in the analysis
- Number included in analyses: 155 at 3 months, 151 at 6 months, 143 at 12 months and 141 at 124 months in the APM group; 158 at 3 months, 146 at 6 months, 136 at 12 months and 141 at 24 months in the PT group. 8/158 (5%) participants in the APM group refused surgery. 17/161 (10.5%) in the PT group did not complete the PT protocol. 35/161 (21.6%) participants in the PT group had APM within 6 months of randomisation.

Inclusion criteria

- Age 45 to 70 years
- Knee pain
- Non-obstructive meniscal tear confirmed by MRI

Exclusion criteria

- Knee locking
- Prior knee surgery
- Instability caused by an anterior or posterior cruciate ligament rupture
- Tumour suspected of malignancy, detectable on MRI
- Severe osteoarthritis (Kellgren-Lawrence score of 4)
- BMI > 35
- American Society of Anesthesiologists (ASA) 4-5 participants
- General disease that affects physical function or systemic medication/abuse of glucocorticoids
- Any other medical condition or treatment interfering with the completion or assessment of the trial; for example, contraindications to MRI or surgery
- Drugs or alcohol abuse
- Participants unable to fill out the Dutch questionnaires
- Associated injuries on the index knee consisting of symptomatic partial or total tear of the anterior cruciate ligament (ACL), posterior cruciate ligament tear, injury to the lateral or posterolateral ligament complex with significant laxity

Baseline characteristics

Arthroscopic partial meniscectomy group

Mean (SD) age: 57.6 (6.5)

No. of male/female: 78/80

Mean (SD) BMI: 26.7 (3.8)

No. (%) mechanical complaints: 56 (35.4)

No. (%) medial meniscal involvement: 126 (79.7)

No. (%) osteoarthritis score (KL classification): 0 = 18 (12.0); 1 = 81 (54.0); 2 = 45 (30.0); 3 = 6 (4.0)

Mean (SD) Knee function International Knee Documentation Committee (IKDC) score (0 = most limitations to 100 = no limitations): 44.8 (16.6)

Median (IQR) Knee pain on VAS (0 = no pain to 100 = worst pain): 61.1 (44.9-83.4)

Physical therapy group

Mean (SD) age: 57.3 (6.8)

No. of male/female: 79/81

Mean (SD) BMI: 27.2 (4.0)

Van de Graaf 2018 *(Continued)*

	No. (%) mechanical complaints: 67 (41.6) No. (%) medial meniscal involvement: 136 (84.5) No. (%) osteoarthritis score (KL classification): 0 = 15 (10.1); 1 = 74 (49.7); 2 = 55 (36.9); 3 = 5 (3.3) Mean (SD) Knee function IKDC score (0 = most limitations to 100 = no limitations): 46.5 (14.6) Median (IQR) Knee pain on VAS (0 = no pain to 100 = worst pain): 59.3 (44.9-77.4) **Pre-treatment group differences:** there are no pre-treatment group differences between the groups.
Interventions	**Arthroscopic partial meniscectomy (APM)** Arthroscopic partial meniscectomy was performed within 4 weeks of randomisation in an outpatient clinic under general or spinal anaesthesia by orthopaedic surgeons experienced in arthroscopic surgery, or orthopaedic residents skilled in arthroscopic surgery under supervision of an orthopaedic surgeon. Standard anteromedial and anterolateral portals were introduced for inspection of the knee joint. The affected meniscus was partially removed until a stable and solid meniscus remained. All participants received perioperative instructions and a home exercise program. Participants were only referred to PT after APM if they did not recover as anticipated as defined by the Dutch Orthopedic Association guidelines. **Physical Therapy (PT)** Participants were referred to PT clinics and their initial PT session was scheduled within 2 weeks after randomisation. Participating PT clinics were instructed about the exercise protocol by a knee-specialised physical therapist or the primary investigator, prior to the first participant's referral. The PT exercise protocol was developed by a knee-specialised physical therapist and consisted of 16 sessions of 30 minutes each conducted over 8 weeks. The PT protocol comprised cardiovascular, coordination/balance, and closed kinetic chain strength exercises (in which the distal part of the extremity is fixed to an object that is stationary). If PT failed, the participant was allowed to attend additional PT sessions or have APM, depending on their preference. **Post-intervention:** both groups received the same home exercise instructions. The home exercise program consisted of one leg standing during 60 seconds and a step-down exercise comprising 3, 9, 10 repetitions, twice a week.
Outcomes	Outcomes were measured at 3, 6, 12 and 24 months **Primary outcome** • Change in knee physical function from baseline to 2 years measured on the International Knee Documentation Committee (IKDC) Subjective Knee Form (scores range from 0 = most limitations to 100 = no limitations in daily and sports activities and the absence of symptoms) **Secondary outcomes** • Knee pain on weight-bearing and at rest measured on VAS (scores range from 0 = no pain to 100 = worst pain) • General health measured using SF-36 (ranging from 0 = worst health to 100 = better health) (data only reported for Physical Component Score) • Quality of life measured using EQ-5D (data not reported) • Radiographic progression of osteoarthritis using KL classification scores (range from 0 = no osteophytes or joint space narrowing indicating no osteoarthritis to 4 ≥ 50% joint space narrowing indicating severe osteoarthritis) • Activity measured on Tegner Activity Scale scores range from 0 = no activity to 10 = higher activity • Physical performance tests (squatting with duck-walk, Thessaly test, McMurray test, range of motion, joint line tenderness, existence of knee joint effusion) • Participant-specific complaints questionnaire • Participant expectation of treatment and their satisfaction (data not reported) • Knee replacement

Van de Graaf 2018 *(Continued)*

- Adverse events: minor, moderate, severe
- Resource utilisation - rehospitalisation, intervention and other healthcare costs, paid help at home, informal care, work absenteeism and presenteeism, unpaid productivity costs
- Function measured using Patient Specific Functional Scale (PSFS)

Outcomes used in this review at 3, 6 and 24 months

- Pain - knee pain on weight-bearing measured on VAS (0 to 100, lower score = less pain)
- Function - change in mean knee function from baseline to 2 years - IKDC scores (0 to 100, higher score = better function)
- Knee replacement
- Serious adverse events: proportion in each group with serious adverse events
- Total adverse events: proportion in each group with any adverse events
- Progression of knee OA (to note: data couldn't be combined in Analysis 7.4 as mean KL scores were reported for each group and not number of those with OA progression)

Notes

Funding: this study was funded by the Netherlands Organization for Health Research and Development (in Dutch: ZonMw; grant 837002009), Zilverenkruis Health Insurance (grant Z436), and the Foundation of Medical Research of the OLVG, Amsterdam (grant 15u.025).

Trial registration: ClinicalTrials.gov number NCT01850719

Adverse events

Arthroscopic partial meniscectomy

Serious adverse events:

No.of events = 9

No. (%) of participants affected = 9/159 (5.7)

Nature of event: neurological events including intracranial malignancy (1), lymph node malignancy (1), rectal polyp (1), knee replacement (2), arthroscopy in affected knee (2), arthroscopy in opposite knee (1), other knee surgery (1)

Other adverse events *:*

No.of events = 9

No. (%) of participants affected = 9 (5.6)

Nature of event: reactive arthritis (1), knee pain resulting in extra consultation (6), Pain in back, hip or foot (2)

Total adverse events:

No. (%): 18/159 (11.32)

Physical therapy

Serious adverse events:

No.of events = 8

No. (%) of participants affected = 8/162 (4.94)

Nature of event: acute myocardial infarction (1), sudden death (1), neurological event (1), alcoholic pancreatitis (1), arthroscopy (1), knee replacement (3)

Other adverse events *:*

No.of events = 4

Van de Graaf 2018 *(Continued)*

No. (%) of participants affected = 4 (2.4)

Nature of event: Knee pain resulting in extra consultation (2) other musculoskeletal (2)

Total adverse events:

No. (%): 12/162 (7.41)

Knee surgery (replacement or osteotomy)

Arthroscopic partial meniscectomy:

No. (%) of participants = 2/159 (1.2)

Other surgery: 3/159 had a re-arthroscopy, 1/159 'other' surgery

Physical therapy:

No. (%) of participants = 3/162 (1.8)

Other surgery: 1/162 arthroscopy

Progression of knee OA: OA severity in the arthroscopy group progressed from 1.3 points at baseline to 1.6 points at 24 months (MD 0.37 points, 95% CI 0.25 to 0.49) and in the physical therapy group from 1.3 points at baseline to 1.5 points at 24 months (MD 0.18 points, 95% CI 0.04 to 0.31). Mixed-model analysis found no significant between-group difference (0.10 points more progression in the arthroscopy group, 95% CI -0.05 to 0.26, P = 0.18)

Withdrawals: two participants (1 from each group) withdrew immediately after randomisation; 18/158 in the APM group and 14/161 in the PT group due to loss to follow-up.

Cross-overs: 47/161 (29%) participants assigned to the exercise group had arthroscopic surgery within two years' follow-up, and 8/159 (5%) participants assigned to the arthroscopy group did not have the procedure

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomisation was performed using a computerised software program (TENALEA Clinical Trial Data Management system) in a 1:1 ratio using random blocks with a maximum block size of 6.
Allocation concealment (selection bias)	Low risk	Randomisation to groups was concealed as it was through an online program.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants, physicians, and physical therapists were not blinded.
Blinding of outcome assessor Self-reported outcomes	High risk	As participants were unable to be blinded due to the nature of the intervention, there could be a risk of bias in the reporting of pain and function.
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	The radiologists assessing X-rays were blinded to treatment allocation so there is low risk of bias in the assessment of osteoarthritis.
Incomplete outcome data (attrition bias)	Low risk	Two participants (1 from each group) withdrew immediately after randomisation without providing a reason. 18/158 (11%) in the APM group and 14/161

Van de Graaf 2018 *(Continued)*

All outcomes		(8.6%) in the PT group were lost to follow-up at 24 months and excluded from the final analysis. Missing outcome data reasonably balanced across groups with similar reasons for missing data across groups
Selective reporting (reporting bias)	Low risk	Trial registered and protocol published. Four deviations from the protocol were clearly outlined in the publication of results - recalculation of SD, inclusion of all time points in the measurement of primary outcome, change to mixed-model analysis from longitudinal analyses and correction of protocol error of 10% loss to follow-up to 20%. Some secondary outcomes listed in the protocol such as resource utilisation, health-related quality of life, patient-specific complaints, participant expectations, and participant satisfaction were not analysed and authors reported they will be analysed and reported separately. The authors reported median and IQR data; however, upon request, they provided mean and SD data which has been used in the analysis
Other bias	Low risk	No other bias apparent

Vermesan 2013

Study characteristics	
Methods	**Study design:** single centre, parallel-group, two-arm, randomised controlled trial **Setting:** hospital in Italy **Trial time period:** not reported **Interventions:** arthroscopic surgery versus a single intra-articular glucocorticoid injection **Sample size calculations:** not reported **Analysis:** not reported
Participants	**Number of participants** • Number of participants screened: 120 • Number of participants at enrolment: 120 • Number randomised: 120 participants: 60 to the arthroscopic debridement group and 60 to the glucocorticoid injection group • Number included in analyses: 120 participants: 60 in the arthroscopic debridement group and 60 in the glucocorticoid injection group (at 1 month); 98 participants: 50 in the arthroscopic debridement group and 48 in the glucocorticoid injection group (at 12 months) **Inclusion criteria** • Non-traumatic symptomatic knees which had degenerative lesions of the medial compartment (cartilage and meniscus) on MRI **Exclusion criteria:** not reported **Baseline characteristics** *Arthroscopic surgery* • Mean (SD) age: 59.2 (7.5) years • No. women: 49 • Mean (SD) body-mass index: 32.7 (6.4) • Mean (SD) onset of symptoms: 3 (1.5) months • Mean (SD) Oxford Knee scores: 29.1 (3.7)

Vermesan 2013 *(Continued)*

	Glucocorticoid injection • Mean (SD) age: 57.6 (7.8) years • No. women: 46 • Mean (SD) body-mass index: 31.9 (6.2) • Mean (SD) onset of symptoms: 3 (1.7) months • Mean (SD) Oxford Knee scores: 30.3 (3.5) **Pre-treatment group differences:** there were no differences in the baseline characteristics between the two groups.
Interventions	**Arthroscopic surgery** Arthroscopic debridement. No description of this procedure was provided. **Intra-articular glucocorticoid injection** A single intra-articular glucocorticoid injection (1 mL of betamethasone in 4 mL of 1% lidocaine) was administered.
Outcomes	Outcome was measured at 1 month and 1 year of follow-up. Oxford Knee Scores which range from: 0 to 19 = severe knee arthritis; 20 to 29 = moderate to severe arthritis; 30 to 39 = mild to moderate arthritis; 40 to 48 = satisfactory joint function (Dawson 1998) **Outcomes used in this review at 1 month and 1 year** • Function measured using the Oxford Knee Score (0 to 48, higher score = better function) • Knee surgery • Knee-related adverse events
Notes	**Funding**: no information on the funding source was provided **Trial registration:** not done **Adverse events:** only knee-related adverse events reports (total of 5 participants had knee replacement); unclear if trial measured other adverse events **Knee surgery (replacement or osteotomy):** a total of 5 participants (4.2%) in both groups had a knee replacement. Per-group data not given. **Progression of knee OA:** not reported **Withdrawals:** 12/60 in the glucocorticoid injection group and 10/60 in the arthroscopic group

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No description of sequence generation process provided
Allocation concealment (selection bias)	Unclear risk	There was no information on allocation concealment.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding of participants and study personnel was not reported. Probably not done

Vermesan 2013 *(Continued)*

Blinding of outcome assessor Self-reported outcomes	High risk	As blinding of participants was not reported and probably was not done, there is likely to be a risk of bias in the measurement of knee pain and function using the Oxford Knee Score.
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	No assessor-reported outcomes were measured in this trial.
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	There was no loss to follow-up for the 1 month follow-up. 12/60 (20%) participants in the glucocorticoid injection group and 10/60 (16%) participants in the arthroscopic group did not have outcome data at the 12 month follow-up. The reasons for loss to follow-up were not reported.
Selective reporting (reporting bias)	High risk	Trial registration not done and protocol not available. The published article had results for one study outcome - the Oxford Knee Scores. Correlation analysis was reported in the methods but this was not reported in the results
Other bias	Low risk	No other bias apparent

Yim 2013

Study characteristics	
Methods	**Study design:** single centre, parallel-group, two-arm, single-blinded, randomised controlled trial **Setting:** Center for Joint Disease, Chonnam National University Hwasun Hospital, Jeonnam, South Korea **Trial time period:** participants were enrolled between January 2007 and July 2009 **Interventions:** arthroscopic surgery plus home exercise versus non-operative care with physical therapy plus home exercise **Sample size calculations:** the sample size was calculated based on Lysholm Knee Score data obtained from 30 prior cases, where the standard deviation was approximately 18. To test the difference in the minimal clinical relevance of 10 between the 2 groups with 80% power and a significance level of $P < 0.05$, these values were estimated with 54 participants in each group **Analysis:** intention-to-treat
Participants	**Number of participants** • Number of participants screened: 162 (49 declined to participate, 5 did not meet inclusion criteria) • Number of participants at enrolment: 108 • Number randomised: 108 participants: 54 to the arthroscopic meniscectomy group and 54 to the non-operative exercise group • Number included in analyses: 102 participants: 50 in the arthroscopic meniscectomy group and 52 to the non-operative exercise group (at 24 months) **Inclusion criteria** • Participants with a degenerative horizontal tear of the posterior horn of the medial meniscus on magnetic resonance imaging (MRI) • Daily knee pain on the medial side with mechanical symptoms affecting daily living activities despite management at a primary clinic during the previous month **Exclusion criteria**

Yim 2013 *(Continued)*

- History of definite trauma
- Previous knee surgery
- Ligament deficiency
- Systemic arthritis
- Osteonecrosis
- Participants showing a marked degenerative change with grade 2, according to the Kellgren-Lawrence classification

Baseline characteristics

Arthroscopic meniscectomy

- Mean (SD) age: 54.9 (10.3) years
- No. men/women: 9/41
- Mean (SD) body-mass index: 25.0 (2.5)
- Mean (SD) mechanical axis: -0.9 (1.3)
- Mean (SD) maximal flexion, deg: 139 (6.9)
- Kellgren–Lawrence grade — no: 0 = 39,1 = 11
- Mean (SD) VAS score: 5.2 (1.8)
- Mean (SD) Lysholm score: 64.0 (11.2)
- Mean (range) duration of symptoms (months): 8.4 (6 weeks-123 months)

Non-operative group

- Mean (SD) age: 57.6 (11.0) years
- No. men/women: 12/40
- Mean (SD) body-mass index: 26.4 (1.9)
- Mean (SD) mechanical axis: -1.1 (1.4)
- Mean (SD) maximal flexion, deg: 141 (10.3)
- Kellgren–Lawrence grade — no: 0 = 35,1 = 17
- Mean (SD) VAS score: 4.9 (1.5)
- Mean (SD) Lysholm score: 65.2 (10.8)
- Mean (range) duration of symptoms (months): 8.2 (2-81 months)

Pre-treatment group differences: there were no differences in baseline characteristics between the two groups.

Interventions	**Arthroscopic surgery** Arthroscopic meniscectomy plus home exercises. Arthroscopic meniscectomy was carried out by a single experienced orthopaedic surgeon, using a 5.5-mm, 30 degree arthroscope and a pressure-controlled irrigation system. The procedure in each case was limited to resection with limited debridement of the articular surface lesion. Participants who underwent additional procedures, such as curettage, abrasion arthroplasty, or subchondral drilling for any articular lesions, were excluded from this study (n = 3). No participant underwent total meniscectomy or peripheral meniscal repair. All participants were discharged on the day after surgery. Subsequently, participants were permitted to use co-interventions, such as analgesics or NSAIDs, within 2 weeks. All participants were then provided with a home exercise program, which was conducted unsupervised, using the same protocol as the non-operative group, for 8 weeks. **Exercise** Non-operative care with physical therapy plus home exercises. All participants in the non-operative group were prescribed drugs, such as analgesics, non-steroidal anti-inflammatory drugs (NSAIDs), or muscle relaxants, depending on clinical symptoms for the first 2 weeks. In addition, they underwent scheduled physical exercise to improve muscle strength, endurance, and flexibility for 60 minutes per session, 3 times weekly, for 3 weeks under the supervision of a physical therapist. After an early, intensive, supervised rehabilitation program to strengthen muscles during the first 3 weeks, all participants

Yim 2013 *(Continued)*

	were provided with a home exercise program, which they conducted unsupervised for 8 weeks. The home exercise program consisted of daily isometric and isotonic muscle exercises. This included: 1. Stretching of knee extensors and flexors (0-8 times per week; 3 times per day; 1min/muscle group); 2. Knee extension in sitting position (0-8 times per week; 3 times per day; 3 x 10 repetitions); 3. Knee flexion in sitting position (0-8 times per week; 3 times per day; 3 x 10 repetitions); 4. Stationary bicycling (0-8 times per week; 3 times per day; gradual increase every 15 minutes); 5. Half squats with < 45 degrees of flexion with weights (5-8 times per week; 3 times per day; 3 x 10 repetitions); 6. Squats with full flexion with weights (5-8 times per week; 3 times per day; 3 x 10 repetitions). Participants were instructed to perform the exercises with some strain but almost pain-free and not adversely influencing the affected knee. **Post-intervention** Participants in both groups were permitted to use co-interventions, such as analgesics, muscle relaxants or NSAIDs, within 2 weeks.
Outcomes	Outcomes were assessed at baseline and at 3, 6, 12 and 24 months follow-up. **Study outcomes** • Mean pain related to specific activities (e.g. stair climbing, squatting, and standing up and sitting down). Measured using 10-point visual analogue scale (VAS) scale (horizontal line that was 10 cm in length) • Pain relief - measured using 10-point VAS scale. Scores categorised as ‘‘complete relief’’ (0 or 1 point on scale), ‘‘improved’’ (> 2-point decrease), or ‘‘persistent’’ (changes within 2 points) • Lysholm Knee Score - scores range from 0 to 100 with lower scores reflecting more severe symptoms • Tegner activity scale - scores range from 0 to 100 with lower scores reflecting more severe symptoms • Participant satisfaction with management, based on knee joint condition and degree of interference with everyday life - rated as ‘‘very satisfied’’ (treatment met participant expectations), ‘‘satisfied’’ (treatment helped, and participant would undergo this treatment option), or ‘‘dissatisfied’’ (participant was the same or worse than before). • Osteoarthritic changes, observed by roentgenography (anteroposterior, lateral, and Merchant views), were graded using the Kellgren-Lawrence classification (grade 1 (doubtful): doubtful joint space narrowing and possible osteophytic lipping; grade 2 (minimal): definite osteophytes and possible joint space narrowing; grade 3 (moderate): moderate multiple osteophytes, definite narrowing of joint space and some sclerosis and possible deformity of bone ends; grade 4 (severe): large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone ends). Participants with grade ≥ 2 arthritis (definite osteophytes or definite narrowing of the joint space on plain radiography) with clear osteophytes were defined as having OA. **Outcomes used in this review at 3, 6 and 24 months** • Pain - mean knee pain during activity, measured on a 10-point VAS scale (endpoint score), lower score = less pain • Function - Lysholm Knee Score, higher score = better outcome with fewer symptoms and disability • Participant-reported treatment success - satisfaction with management at 24 months, measured as number of participants reporting 'very satisfied' or 'satisfied' • Progression of knee OA
Notes	**Funding:** not reported **Trial registration:** not registered **Adverse events:** unclear if measured; not reported

Yim 2013 *(Continued)*

Knee surgery (replacement or osteotomy): not reported

Progression of knee OA: 2/50 (4.0%) participants in the meniscectomy group and 3/52 (5.8%) in the non-operative group at 2-year follow-up

Withdrawals: 4/54 (7.4%) in the meniscectomy group and 2/54 (3.7%) in the non-operative group

Data imputations: SDs for VAS pain and Lysholm Knee Function Score not reported at follow-up and not provided by authors on request. We used Yim 2013 baseline SDs in analyses Analysis 2.1; Analysis 2.2; Analysis 13.1; and Analysis 13.2.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No description of sequence generation process provided
Allocation concealment (selection bias)	Unclear risk	Allocation concealment unclear. "Closed-envelope technique" reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding of participants and study personnel not done
Blinding of outcome assessor Self-reported outcomes	High risk	Participants were aware of treatment allocation; thus, there is risk of bias in measurements of pain, knee function and activity and participant satisfaction with treatment
Blinding of outcome assessor Assessor-reported outcome (knee replacement)	Low risk	Outcome assessors for progression of knee OA probably blinded: "Clinical outcome measures and physical examinations were conducted by independent authors not involved in the treatment"
Incomplete outcome data (attrition bias) All outcomes	Low risk	4/54 (7.4%) (not meeting inclusion criteria = 3, loss to follow-up = 1) in the arthroscopic meniscectomy group and 2/54 (3.7%) (cross-over to the other group = 1, loss to follow-up = 1) in the non-operative exercise group were excluded from the analysis at 24 months
Selective reporting (reporting bias)	High risk	Trial registration not done and protocol not available. The authors reported mean and range for study outcomes but standard deviations and confidence intervals were not reported. Adverse events were not reported
Other bias	Unclear risk	An unspecified number of participants in the arthroscopic surgery group were not prescribed exercise

ACR: American College of Rheumatology; **ADL:** activities of daily living; **AIMS:** Arthritis Impact Measurement Scale; **AIMS2-P:** pain subscale of the Arthritis Impact Measurement Scale; **BMI:** body mass index; **cc:** cubic centimetres; **DVT:** deep vein (venous) thrombosis; **EQ-5D 3L:** EuroQoL 5-dimension 3-level quality of life questionnaire; **IQR:** interquartile range; **ITT:** intention-to-treat; **KL grade:** Kellgren-Lawrence classification grade; **KOOS:** Knee injury and Osteoarthritis Outcome Score;**KOOS 4/5:** derived from 4 or 5 KOOS subscale scores; **MI:** myocardial infarction; **MRI:** magnetic resonance imaging; **NSAIDS:** non-steroidal anti-inflammatory drugs; **OA:** osteoarthritis; **OARSI:** Osteoarthritis Research Society International; **PT:** physical therapy; **QoL/QOL:** quality of life; **ROM:** range of motion; **SD:** standard deviation; **SMD:** standardised mean difference; **VAS:** visual analogue scale; **WOMAC:** Western Ontario and McMaster Universities Osteoarthritis Index; **WOMET:** Western Ontario Meniscal Evaluation Tool

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Ahn 2015	Study population did not have degenerative knee disease
Biedert 2000	Study population did not have degenerative knee disease
Bisson 2015	Study population did not have degenerative knee disease
Bradley 2002	Intervention was not arthroscopic surgery
Hubbard 1996	Study population did not have degenerative knee disease or osteoarthritis
Kalunian 2000	Intervention was not arthroscopic surgery
Lee 2020	Not an RCT
Lu 2018	Not an RCT
Ma 2020	Not an RCT
Marsh 2016	Study examined cost-effectiveness of arthroscopic surgery
Pan 2020	Not an RCT
Rimington 2009	Not an RCT
Wijn 2020	Not an RCT
Zhang 2018	Intervention was arthroscopic joint lavage
Zhao 2018	Not an RCT

Characteristics of studies awaiting classification *[ordered by study ID]*

Kang 2005

Methods	
Participants	
Interventions	
Outcomes	
Notes	Article is in Chinese awaiting translation into English

NCT00562822

Methods	**Study design:** prospective single-blind, single-centre, parallel, two-arm randomised controlled trial **Setting:** London Health Sciences Centre - University Hospital, London, Ontario, Canada **Trial time period:** January 1999 to August 2007

NCT00562822 *(Continued)*

	Interventions: arthroscopic surgery plus optimised physical and medical therapy versus optimised physical and medical therapy **Sample size calculations:** estimated enrolment of 188 participants, no sample size calculations reported **Analysis:** not reported
Participants	**Inclusion Criteria** Idiopathic or secondary osteoarthritis of the knee with Grade 2 to 4 radiographic severity as defined by the modified Kellgren-Lawrence classification **Exclusion Criteria** • Participants with inflammatory or post-infectious arthritis • Those who had undergone previous arthroscopic treatment for knee osteoarthritis • Those with isolate Grade III to IV medical compartment osteoarthritis with greater than 5 degrees of varus deformity
Interventions	**Arthroscopic surgery plus optimised physical and medical therapy:** arthroscopic surgery to treat unresolved symptoms of osteoarthritis of the knee **Optimised physical and medical therapy:** treatment with physical and medical therapy alone
Outcomes	**Primary outcome measures:** Function, pain and quality of life based on the WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index) scores (time frame: 2 years) **Secondary outcome measures:** Health utility using the standard-gamble technique (time frame: 24 months)
Notes	**Clinical trial registration:** NCT00562822 **Trial status:** recruitment completed **Estimated study completion date:** recruitment completed in August 2007; however, results not available yet

Characteristics of ongoing studies *[ordered by study ID]*

NCT02113280

Study name	DEMAND - DEgenerative Meniscal Tears - Arthroscopy vs. Dedicated Exercise **Official title:** Randomised Controlled Trial Comparing Arthroscopy With Physiotherapy for Degenerative Meniscal Tears
Methods	**Study design:** prospective parallel-arm randomised controlled trial **Setting:** North Tyneside General Hospital, UK **Trial time period:** December 2015 to December 2018 **Interventions:** arthroscopic surgery versus physiotherapy **Sample size calculations:** not reported **Analysis:** not reported

NCT02113280 *(Continued)*

Participants	**Inclusion criteria** • Age over 45 years • Knee pain in the presence of a medial meniscal tear on MRI **Exclusion criteria** • History of trauma to the knee or ipsilateral lower limb in the past 2 years • Inability to engage in post-operative rehabilitation • Lacking capacity to consent • Evidence of infection • Previous knee surgery other than arthroscopy (diagnostic or partial meniscectomy) • Neurological disease • Inflammatory arthritis • Loose bodies • Ligament injuries causing symptomatic instability • Women who are pregnant • Have current signs or symptoms of severe, progressive or uncontrolled renal, hepatic, haematological, gastrointestinal, endocrine, pulmonary, cardiac, neurologic, or cerebral disease • Uncontrolled disease states, such as moderate/severe asthma, chronic obstructive pulmonary disease (COPD) or inflammatory bowel disease, where flares are commonly treated with oral or parenteral glucocorticoids, or recurrent infections
Interventions	**Arthroscopy:** participants to receive knee arthroscopy and meniscal debridement **Physiotherapy:** outpatient standardised physiotherapy regime with focus on exercise therapy
Outcomes	Outcomes will be measured at baseline, 6 weeks and 6, 12 and 24 months. **Primary outcome** • Change in Knee injury and Osteoarthritis Outcome Score **Secondary outcomes** • SF-12 (12-item short form survey) • Pain visual analogue score
Starting date	December 2015
Contact information	Sarah Johnson-Lynn, MBBS Specialty Registrar, Northumbria Healthcare NHS Foundation Trust 08448118111 ext 2508 selynn@doctors.org.uk Derek Kramer, MBBS 08448118111 ext 2508 derek.kramer@northumbria-healthcare.nhs.uk
Notes	**Clinical trial registration:** NCT02113280 **Trial status**: withdrawn (failed funding application) **Expected completion date**: December 2018

NCT02995551

Study name	Danish RCT on Exercise versus Arthroscopic Meniscal Surgery for Young Adults (DREAM) **Official title:** Danish Rct on Exercise versus Arthroscopic Meniscal Surgery for Young Adults (DREAM) - A Randomized Controlled Trial of Meniscal Tear Treatment in Young Adults

NCT02995551 *(Continued)*

Methods	**Study design:** prospective, parallel-arm, multicentre, randomised controlled trial **Setting:** Denmark **Trial time period:** January 2017 to June 2018 **Interventions:** arthroscopic meniscal surgery versus individualised supervised exercise therapy and education **Sample size calculations:** 59 participants in each of the intervention groups is needed (assuming a common standard deviation (SD) of 16.5, power = 90%, alpha level = 0.05) to detect a clinically relevant difference of 10 points in the primary outcome ($KOOS_4$) from baseline to 12 months' follow-up. A total of 140 participants will be recruited to account for loss to follow-up (19%). **Analysis:** intention-to-treat
Participants	**Inclusion criteria** • Adults aged 18 to 40 years with knee pain • Clinical history and symptoms consistent with meniscal tear and meniscal tear verified on magnetic resonance imaging (MRI) • Deemed eligible for meniscal surgery (i.e. repair or resection) by the examining orthopaedic surgeon • Willing to participate in 12 weeks of supervised exercise twice a week and undergo surgery for the meniscal tear as soon as possible **Exclusion criteria** • Previous knee surgery on the affected knee • Clinical suspicion (acute locking of knee and/or extension deficit) of displaced 'bucket handle' tear confirmed by MRI • Fracture of the affected extremity within the previous 6 months • Complete rupture of one or more knee ligaments • Participation in supervised systematic exercise for knee problems within the last 3 months prior to recruitment • Other reasons for exclusion (unable to understand Danish, mentally unable to participate, etc).
Interventions	**Arthroscopic meniscal surgery** Arthroscopic meniscal repair or resection will be conducted at the discretion of the operating surgeon at one of the six hospitals. The specific surgical procedure (i.e. repair or resection) cannot be determined before the surgeon has visual confirmation about the exact knee pathology and extent of the meniscal tear at arthroscopy. **Exercise therapy and patient education** Participants allocated to exercise therapy and participant education will twice weekly participate in a 12-week individualised, supervised exercise programme (approximately 60 to 90 minutes/session) tailored to 18 to 40 years old participants with meniscal tear
Outcomes	Outcomes will be measured at 3, 6 and 12 months' follow-up **Primary outcome** Change in Knee injury and Osteoarthritis Outcome Score (KOOS) scores from baseline to follow-up. $KOOS_4$ is the mean score for the KOOS subscales pain, symptoms, function in sports and recreational activities (Sport/Rec) and quality of life (QOL) **Secondary outcomes:** • KOOS subscales individual scores

NCT02995551 *(Continued)*

- Western Ontario Meniscal Evaluation Tool (WOMET) scores
- Physical performance:
  - Isometric muscle strength using the Fysiometer
  - Maximum knee-bends in 30 seconds
  - One-leg hop for distance
  - 6 metre timed hop
- Adverse events

Starting date	January 2017
Contact information	Principal Investigator: Søren Thorgaard Skou, PT, PhD, Assistant Professor, University of Southern Denmark Tel: +4523708640; Email: stskou@health.sdu.dk
Notes	**Clinical trial registration:** NCT02995551 **Trial status:** recruiting participants **Expected completion date**: December 2020

NCT04313569

Study name	Arthroscopic Versus Conservative Treatment of Degenerative Meniscal Tear in Middle Aged Patients in Regard to Pain & Knee Function **Official title:** Arthroscopic Versus Conservative Treatment of Degenerative Meniscal Tear in Middle Aged Patients in Regard to Pain & Knee Function: Comparative Study
Methods	**Study design:** single-centre, parallel-group, two-arm, randomised controlled trial **Setting:** Erbil teaching hospital, Iran **Trial time period:** August 2017 to September 2019 **Interventions:** arthroscopic meniscectomy versus conservative treatment **Sample size calculations:** 60 participants, 30 in each group **Analysis:** not reported
Participants	**Inclusion criteria** • Age ranging between 40 and 60 years • All were clinically diagnosed to have degenerative medial meniscal tear then confirmed by magnetic resonance imaging (MRI) • Atraumatic continuous pain in medial aspect of knee affecting daily activities, for more than one month despite the treatment of general physician **Exclusion criteria** • Meniscal tear due to trauma • Any rheumatologic knee disease • MRI showing ligament injury, loose bodies, tumours and/or osteochondral defects • Former surgery of knee and lower limb fractures in the last year • Knee joints with osteoarthritis graded 2 or more according to Kellgren-Lawrence scale (Kellgren 1957) on weight-bearing knee X-ray • Lateral meniscus tear

NCT04313569 *(Continued)*

Interventions	**Intervention** - arthroscopic meniscectomy **Control** - conservative treatment Use of medications, like analgesics, muscle relaxants, non-steroidal anti-inflammatory drugs (NSAID) and local painkillers, depending on participants' condition, and physiotherapy, lifestyle and daily activity modification, participant education about positioning of the knee.
Outcomes	**Primary outcomes:** Lysholm Knee Scoring Scale at 1 year: Lysholm Knee score scale has 8 parts (swelling, pain, squatting, support, stair climbing, limping, locking and instability). Score of 100 means no problems in knee. Visual analogue score at 1 year: 10-point Visual Analogue Score scale used in order to measure the severity of knee pain during the study. 0 = no pain; 10 = most severe pain **Secondary outcomes:** None specified
Starting date	1 August 2017
Contact information	Sherwan Ahmed Ali Hamawandi, Assistant Professor of Orthopedic Surgery, Hawler Medical University
Notes	**Clinical trial registration:** NCT04313569 **Trial status:** recruitment completed **Expected completion date:** 30 September 2019. No results available.

NCT04837456

Study name	Metabolic Syndrome and Degenerate Meniscus Tears
Methods	**Study design:** single centre, double-blind, parallel-group, four-arm randomised controlled trial **Setting:** First Affiliated Hospital of Jinzhou Medical University, China **Trial time period:** June 2017 to March 2020 **Interventions:** calorie-restricted diet and exercise intervention group; libitum diet and waiting list control group; early arthroscopic partial meniscectomy (APM) (syndrome within 3 to 6 months) group or a delayed APM (syndrome more than 6 months) group **Sample size calculations:** 180 participants **Analysis:** not reported
Participants	**Inclusion criteria** • Age between 35 and 70 years old • Clinical diagnosis of metabolic syndrome • Clinical diagnosis of 3 grade degeneration meniscus lesions **Exclusion criteria** • Acute knee injury such as car crash or acute sports injury • History of knee surgeries

NCT04837456 *(Continued)*

	• Rheumatoid arthritis or serious knee osteoarthritis with deformity • Contraindications to magnetic resonance imaging (MRI) • Severe cardiopulmonary disease • Musculoskeletal or neuromuscular impairments • Poor visual, hearing, or cognitive function
Interventions	**Calorie restricted diet and exercise intervention:** A balanced diet that provided an energy deficit of 800 kcal/day from their daily energy requirement. Macronutrient content of low caloric diet, expressed as percentage of ingested energy with carbohydrates 45-65%; fat 20-35%; and protein 10-35%. Each session was approximately 150 minutes one week for six months and consisted of aerobic exercises, resistance training, and exercises to improve flexibility and balance. **Libitum diet and waiting list control group:** participants then underwent a calorie of 2000 calorie above based on libitum free diets recommended to adults and normal physical activity without exercise during the program. **Early APM group:** Early APM group participants received APM with symptoms within 3 to 6 months. **Delayed APM group:** Delayed APM group recruit participants with symptoms lasting for more than 6 months.
Outcomes	**Primary outcomes** Knee KOOS 4 WOMAC International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form The WOMET score Kellgren-Lawrence grade BMI Lysholm knee score Blood tests
Starting date	1 June 2017
Contact information	Hongyu Wang, chief resident, The First People's Hospital of Jingzhou
Notes	**Clinical trial registration:** NCT04837456 **Trial status:** recruitment completed **Expected completion date:** 1 March 2021

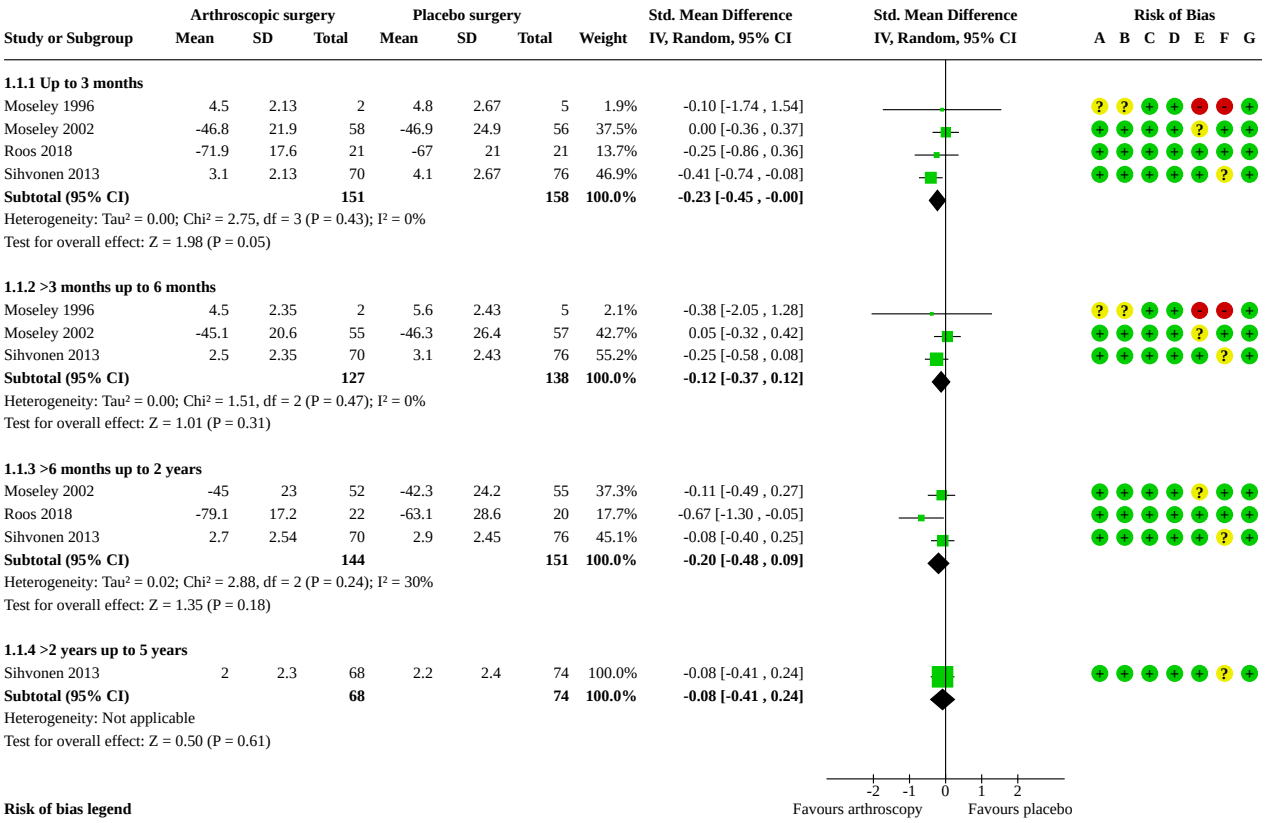
DATA AND ANALYSES

Comparison 1. Arthroscopic surgery versus placebo surgery

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Pain (lower score=less pain)	4		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1.1 Up to 3 months	4	309	Std. Mean Difference (IV, Random, 95% CI)	-0.23 [-0.45, -0.00]
1.1.2 >3 months up to 6 months	3	265	Std. Mean Difference (IV, Random, 95% CI)	-0.12 [-0.37, 0.12]
1.1.3 >6 months up to 2 years	3	295	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.48, 0.09]
1.1.4 >2 years up to 5 years	1	142	Std. Mean Difference (IV, Random, 95% CI)	-0.08 [-0.41, 0.24]
1.2 Function (higher score=better function)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.2.1 Up to 3 months	3	302	Std. Mean Difference (IV, Random, 95% CI)	0.01 [-0.22, 0.23]
1.2.2 >3 months up to 6 months	2	257	Std. Mean Difference (IV, Random, 95% CI)	0.05 [-0.20, 0.29]
1.2.3 >6 months up to 2 years	3	293	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.27, 0.47]
1.2.4 > 2 years up to 5 years	1	142	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.48, 0.18]
1.3 Knee-specific quality of life (higher score=better)	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.3.1 Up to 3 months	2	188	Std. Mean Difference (IV, Random, 95% CI)	0.31 [0.02, 0.59]
1.3.2 >3 months up to 6 months	1	146	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.20, 0.45]
1.3.3 >6 months up to 2 years	2	188	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.24, 0.70]
1.3.4 >2 years up to 5 years	1	142	Std. Mean Difference (IV, Random, 95% CI)	-0.02 [-0.35, 0.31]
1.4 Generic quality of life (higher score=better)	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.4.1 Up to 3 months	1	42	Std. Mean Difference (IV, Random, 95% CI)	-0.56 [-1.18, 0.06]
1.4.2 >6 months up to 2 years	2	188	Std. Mean Difference (IV, Random, 95% CI)	0.15 [-0.28, 0.58]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.5 Participant-reported success	3	189	Risk Ratio (M-H, Random, 95% CI)	1.11 [0.66, 1.86]

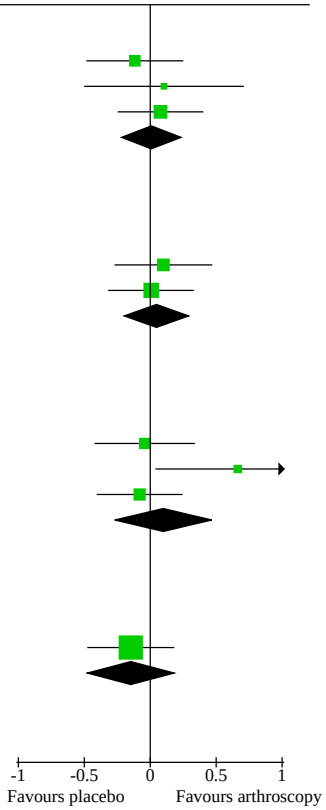
Analysis 1.1. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 1: Pain (lower score=less pain)



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessor: Self-reported outcomes
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

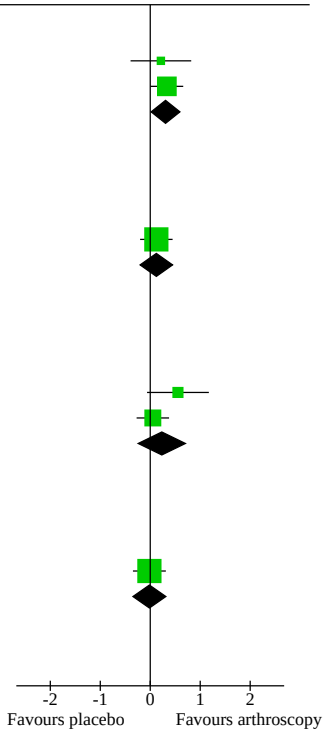
Analysis 1.2. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 2: Function (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo surgery Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.2.1 Up to 3 months								
Moseley 2002	49.6	24.2	58	52.4	23.5	56	37.8%	-0.12 [-0.48 , 0.25]
Roos 2018	76.7	15.7	21	74.8	19.8	21	13.9%	0.10 [-0.50 , 0.71]
Sihvonen 2013	77.2	15.79	70	75.9	17.35	76	48.3%	0.08 [-0.25 , 0.40]
Subtotal (95% CI)			**149**			**153**	**100.0%**	**0.01 [-0.22 , 0.23]**
Heterogeneity: Tau² = 0.00; Chi² = 0.72, df = 2 (P = 0.70); I² = 0%								
Test for overall effect: Z = 0.07 (P = 0.94)								
1.2.2 >3 months up to 6 months								
Moseley 2002	51	25.9	55	48.4	25.9	57	43.6%	0.10 [-0.27 , 0.47]
Sihvonen 2013	82.8	16.01	70	82.7	14.58	75	56.4%	0.01 [-0.32 , 0.33]
Subtotal (95% CI)			**125**			**132**	**100.0%**	**0.05 [-0.20 , 0.29]**
Heterogeneity: Tau² = 0.00; Chi² = 0.14, df = 1 (P = 0.71); I² = 0%								
Test for overall effect: Z = 0.38 (P = 0.71)								
1.2.3 >6 months up to 2 years								
Moseley 2002	47.9	26.6	52	49	27.2	54	36.8%	-0.04 [-0.42 , 0.34]
Roos 2018	85.1	15.5	22	71.2	25	20	22.1%	0.66 [0.04 , 1.29]
Sihvonen 2013	82.2	15.89	69	83.4	13.79	76	41.1%	-0.08 [-0.41 , 0.25]
Subtotal (95% CI)			**143**			**150**	**100.0%**	**0.10 [-0.27 , 0.47]**
Heterogeneity: Tau² = 0.06; Chi² = 4.53, df = 2 (P = 0.10); I² = 56%								
Test for overall effect: Z = 0.53 (P = 0.60)								
1.2.4 > 2 years up to 5 years								
Sihvonen 2013	83.7	14.3	68	85.8	14	74	100.0%	-0.15 [-0.48 , 0.18]
Subtotal (95% CI)			**68**			**74**	**100.0%**	**-0.15 [-0.48 , 0.18]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.88 (P = 0.38)								
Test for subgroup differences: Chi² = 1.19, df = 3 (P = 0.76), I² = 0%								



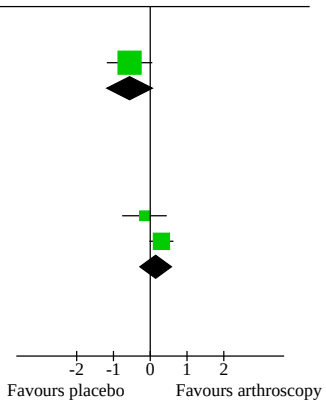
Analysis 1.3. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 3: Knee-specific quality of life (higher score=better)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo surgery Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.3.1 Up to 3 months								
Roos 2018	50.1	16.7	21	46	20.8	21	22.5%	0.21 [-0.39 , 0.82]
Sihvonen 2013	76.8	17.93	70	69.7	23.8	76	77.5%	0.33 [0.01 , 0.66]
Subtotal (95% CI)			**91**			**97**	**100.0%**	**0.31 [0.02 , 0.59]**
Heterogeneity: Tau² = 0.00; Chi² = 0.12, df = 1 (P = 0.73); I² = 0%								
Test for overall effect: Z = 2.09 (P = 0.04)								
1.3.2 >3 months up to 6 months								
Sihvonen 2013	81.7	21.13	70	79.1	21.21	76	100.0%	0.12 [-0.20 , 0.45]
Subtotal (95% CI)			**70**			**76**	**100.0%**	**0.12 [-0.20 , 0.45]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.74 (P = 0.46)								
1.3.3 >6 months up to 2 years								
Roos 2018	63.5	21.1	22	49.4	28.6	20	35.7%	0.55 [-0.06 , 1.17]
Sihvonen 2013	81	20.77	70	79.9	21.35	76	64.3%	0.05 [-0.27 , 0.38]
Subtotal (95% CI)			**92**			**96**	**100.0%**	**0.23 [-0.24 , 0.70]**
Heterogeneity: Tau² = 0.06; Chi² = 1.99, df = 1 (P = 0.16); I² = 50%								
Test for overall effect: Z = 0.96 (P = 0.34)								
1.3.4 >2 years up to 5 years								
Sihvonen 2013	84.3	17.8	68	84.6	17.9	74	100.0%	-0.02 [-0.35 , 0.31]
Subtotal (95% CI)			**68**			**74**	**100.0%**	**-0.02 [-0.35 , 0.31]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.10 (P = 0.92)								

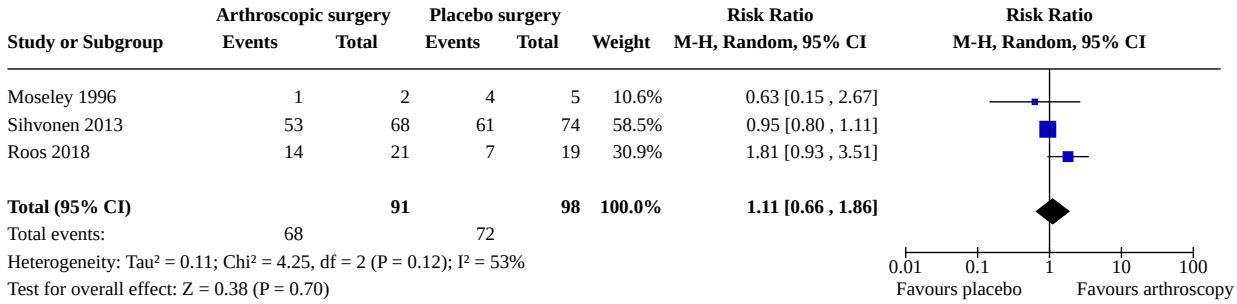


Analysis 1.4. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 4: Generic quality of life (higher score=better)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo surgery Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.4.1 Up to 3 months								
Roos 2018	57.8	6.5	21	61.3	5.7	21	100.0%	-0.56 [-1.18 , 0.06]
Subtotal (95% CI)			**21**			**21**	**100.0%**	**-0.56 [-1.18 , 0.06]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.78 (P = 0.07)								
1.4.2 >6 months up to 2 years								
Roos 2018	55.7	6.9	22	57	9.3	20	34.0%	-0.16 [-0.76 , 0.45]
Sihvonen 2013	0.94	0.06	70	0.92	0.07	76	66.0%	0.30 [-0.02 , 0.63]
Subtotal (95% CI)			**92**			**96**	**100.0%**	**0.15 [-0.28 , 0.58]**
Heterogeneity: Tau² = 0.04; Chi² = 1.72, df = 1 (P = 0.19); I² = 42%								
Test for overall effect: Z = 0.67 (P = 0.50)								



Analysis 1.5. Comparison 1: Arthroscopic surgery versus placebo surgery, Outcome 5: Participant-reported success



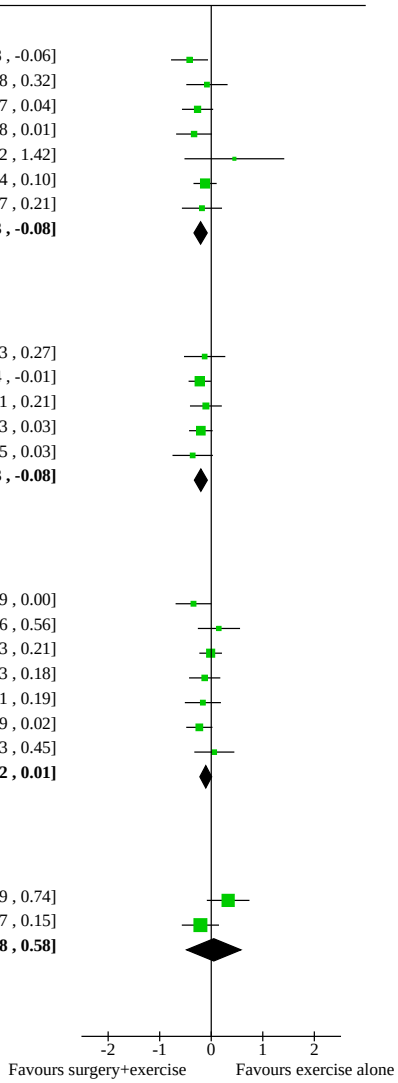
Comparison 2. Arthroscopic surgery versus exercise

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Pain (lower score=less pain)	8		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1.1 Up to 3 months	7	942	Std. Mean Difference (IV, Random, 95% CI)	-0.21 [-0.33, -0.08]
2.1.2 >3 months up to 6 months	5	987	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.33, -0.08]
2.1.3 >6 months up to 2 years	7	1178	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.22, 0.01]
2.1.4 >2 years up to 5 years	2	219	Std. Mean Difference (IV, Random, 95% CI)	0.05 [-0.48, 0.58]
2.2 Function (higher score=better function)	8		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.2.1 Up to 3 months	7	949	Std. Mean Difference (IV, Random, 95% CI)	0.13 [0.00, 0.26]
2.2.2 >3 months up to 6 months	5	988	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.01, 0.24]
2.2.3 >6 months up to 2 years	7	1228	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.01, 0.21]
2.2.4 >2 years up to 5 years	2	219	Std. Mean Difference (IV, Random, 95% CI)	-0.06 [-0.34, 0.23]
2.3 Knee-specific quality of life (higher score=better)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
2.3.1 Up to 3 months	3	347	Mean Difference (IV, Random, 95% CI)	6.87 [2.55, 11.19]
2.3.2 >3 months up to 6 months	1	96	Mean Difference (IV, Random, 95% CI)	0.49 [-8.28, 9.26]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.3.3 >6 months up to 2 years	3	348	Mean Difference (IV, Random, 95% CI)	4.47 [-1.33, 10.28]
2.3.4 >2 years up to 5 years	2	220	Mean Difference (IV, Random, 95% CI)	2.13 [-5.74, 10.00]
2.4 Generic quality of life (higher score=better)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.4.1 Up to 3 months	2	290	Std. Mean Difference (IV, Random, 95% CI)	0.09 [-0.14, 0.32]
2.4.2 >3 months up to 6 months	1	163	Std. Mean Difference (IV, Random, 95% CI)	0.14 [-0.17, 0.45]
2.4.3 >6 months up to 2 years	3	425	Std. Mean Difference (IV, Random, 95% CI)	0.03 [-0.16, 0.22]
2.4.4 >2 years up to 5 years	1	101	Std. Mean Difference (IV, Random, 95% CI)	-0.26 [-0.67, 0.15]
2.5 Participant-reported success	3	532	Risk Ratio (M-H, Random, 95% CI)	1.17 [0.86, 1.59]

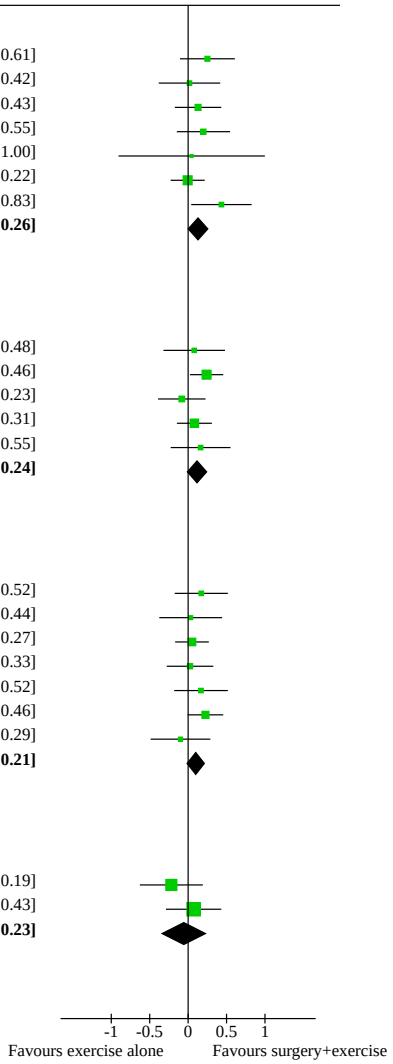
Analysis 2.1. Comparison 2: Arthroscopic surgery versus exercise, Outcome 1: Pain (lower score=less pain)

Study or Subgroup	Arthroscopy plus exercise Mean	SD	Total	Exercise alone Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Up to 3 months								
Gauffin 2014	-77	16.6	66	-69	21.2	57	12.8%	-0.42 [-0.78 , -0.06]
Herrlin 2007	-81.9	16.2	47	-80.6	14.8	49	10.3%	-0.08 [-0.48 , 0.32]
Kirkley 2008	141	109	90	172	124	80	18.0%	-0.27 [-0.57 , 0.04]
Kise 2016	-81.94	16.87	64	-75.64	20.53	65	13.6%	-0.33 [-0.68 , 0.01]
Osteras 2012	2.6	1.1	8	2	1.4	9	1.8%	0.45 [-0.52 , 1.42]
Van de Graaf 2018	30.4	24.6	154	33.4	25.4	151	32.6%	-0.12 [-0.34 , 0.10]
Yim 2013	2.4	1.8	50	2.7	1.5	52	10.9%	-0.18 [-0.57 , 0.21]
Subtotal (95% CI)			**479**			**463**	**100.0%**	**-0.21 [-0.33 , -0.08]**
Heterogeneity: Tau² = 0.00; Chi² = 4.75, df = 6 (P = 0.58); I² = 0%								
Test for overall effect: Z = 3.15 (P = 0.002)								
2.1.2 >3 months up to 6 months								
Herrlin 2007	-83.7	15.7	47	-81.6	16.9	49	9.8%	-0.13 [-0.53 , 0.27]
Katz 2013	21.1	18.1	161	25.2	18.6	169	33.5%	-0.22 [-0.44 , -0.01]
Kirkley 2008	143	113	90	155	118	73	16.4%	-0.10 [-0.41 , 0.21]
Van de Graaf 2018	25.4	27.7	151	31	27.7	145	30.1%	-0.20 [-0.43 , 0.03]
Yim 2013	1.5	1.8	50	2.1	1.5	52	10.2%	-0.36 [-0.75 , 0.03]
Subtotal (95% CI)			**499**			**488**	**100.0%**	**-0.20 [-0.33 , -0.08]**
Heterogeneity: Tau² = 0.00; Chi² = 1.18, df = 4 (P = 0.88); I² = 0%								
Test for overall effect: Z = 3.15 (P = 0.002)								
2.1.3 >6 months up to 2 years								
Gauffin 2014	-84	14.9	70	-78	19.8	60	10.9%	-0.34 [-0.69 , 0.00]
Herrlin 2007	-87.8	16.4	46	-90.1	14	46	7.8%	0.15 [-0.26 , 0.56]
Katz 2013	19.1	17.5	156	19.3	17.6	164	27.3%	-0.01 [-0.23 , 0.21]
Kirkley 2008	168	134	88	185	132	80	14.3%	-0.13 [-0.43 , 0.18]
Kise 2016	-89.37	16	64	-86.69	16.84	62	10.7%	-0.16 [-0.51 , 0.19]
Van de Graaf 2018	19.6	23.8	115	25.5	26.9	125	20.3%	-0.23 [-0.49 , 0.02]
Yim 2013	1.8	1.8	50	1.7	1.5	52	8.7%	0.06 [-0.33 , 0.45]
Subtotal (95% CI)			**589**			**589**	**100.0%**	**-0.11 [-0.22 , 0.01]**
Heterogeneity: Tau² = 0.00; Chi² = 5.77, df = 6 (P = 0.45); I² = 0%								
Test for overall effect: Z = 1.81 (P = 0.07)								
2.1.4 >2 years up to 5 years								
Gauffin 2014	-79.3	20.97	65	-86	18.88	35	48.2%	0.33 [-0.09 , 0.74]
Kise 2016	-90	14.86	62	-86.7	16.37	57	51.8%	-0.21 [-0.57 , 0.15]
Subtotal (95% CI)			**127**			**92**	**100.0%**	**0.05 [-0.48 , 0.58]**
Heterogeneity: Tau² = 0.11; Chi² = 3.70, df = 1 (P = 0.05); I² = 73%								
Test for overall effect: Z = 0.18 (P = 0.86)								



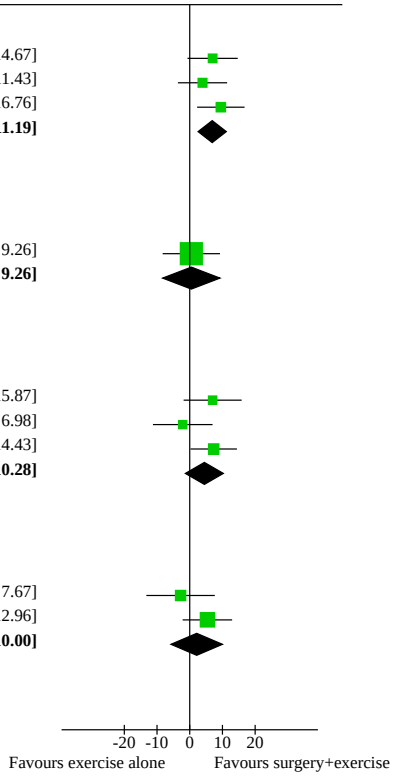
Analysis 2.2. Comparison 2: Arthroscopic surgery versus exercise, Outcome 2: Function (higher score=better function)

Study or Subgroup	Arthroscopy plus exercise Mean	SD	Total	Exercise alone Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 Up to 3 months								
Gauffin 2014	81	18.5	65	76	21.2	57	12.8%	0.25 [-0.11 , 0.61]
Herrlin 2007	86.51	16.85	47	86.22	15.93	49	10.2%	0.02 [-0.38 , 0.42]
Kirkley 2008	-522	341	90	-568	369	80	17.9%	0.13 [-0.17 , 0.43]
Kise 2016	88.74	16.18	64	85.25	18.55	65	13.6%	0.20 [-0.15 , 0.55]
Osteras 2012	40.9	23.1	8	39.7	25.9	9	1.8%	0.05 [-0.91 , 1.00]
Van de Graaf 2018	59.9	16.6	155	60	17.4	158	33.2%	-0.01 [-0.23 , 0.22]
Yim 2013	85.2	11.2	50	80.4	10.8	52	10.6%	0.43 [0.04 , 0.83]
Subtotal (95% CI)			**479**			**470**	**100.0%**	**0.13 [0.00 , 0.26]**
Heterogeneity: Tau² = 0.00; Chi² = 4.66, df = 6 (P = 0.59); I² = 0%								
Test for overall effect: Z = 1.98 (P = 0.05)								
2.2.2 >3 months up to 6 months								
Herrlin 2007	88.4	16.13	47	87.06	17.33	49	9.8%	0.08 [-0.32 , 0.48]
Katz 2013	-14.7	17.8	161	-19	17.9	169	33.3%	0.24 [0.02 , 0.46]
Kirkley 2008	-551	382	90	-520	368	73	16.4%	-0.08 [-0.39 , 0.23]
Van de Graaf 2018	64.7	19.2	151	63.2	17.5	146	30.2%	0.08 [-0.15 , 0.31]
Yim 2013	84.1	11.2	50	82.3	10.8	52	10.3%	0.16 [-0.23 , 0.55]
Subtotal (95% CI)			**499**			**489**	**100.0%**	**0.12 [-0.01 , 0.24]**
Heterogeneity: Tau² = 0.00; Chi² = 3.02, df = 4 (P = 0.55); I² = 0%								
Test for overall effect: Z = 1.81 (P = 0.07)								
2.2.3 >6 months up to 2 years								
Gauffin 2014	86	17.1	70	83	17.8	60	10.5%	0.17 [-0.17 , 0.52]
Herrlin 2007	91.63	14.73	46	91.13	15.02	46	7.5%	0.03 [-0.38 , 0.44]
Katz 2013	-13.7	15.9	156	-14.5	16	164	26.1%	0.05 [-0.17 , 0.27]
Kirkley 2008	-612	448	88	-623	439	80	13.7%	0.02 [-0.28 , 0.33]
Kise 2016	93.53	15.72	65	90.96	14.76	62	10.3%	0.17 [-0.18 , 0.52]
Van de Graaf 2018	71.5	16.4	141	67.7	17.2	148	23.5%	0.23 [-0.01 , 0.46]
Yim 2013	83.2	11.2	50	84.3	10.8	52	8.3%	-0.10 [-0.49 , 0.29]
Subtotal (95% CI)			**616**			**612**	**100.0%**	**0.10 [-0.01 , 0.21]**
Heterogeneity: Tau² = 0.00; Chi² = 2.98, df = 6 (P = 0.81); I² = 0%								
Test for overall effect: Z = 1.73 (P = 0.08)								
2.2.4 >2 years up to 5 years								
Gauffin 2014	83.9	18.78	64	87.9	16.99	36	44.1%	-0.22 [-0.63 , 0.19]
Kise 2016	93	14.66	62	92	13.1	57	55.9%	0.07 [-0.29 , 0.43]
Subtotal (95% CI)			**126**			**93**	**100.0%**	**-0.06 [-0.34 , 0.23]**
Heterogeneity: Tau² = 0.00; Chi² = 1.09, df = 1 (P = 0.30); I² = 8%								
Test for overall effect: Z = 0.39 (P = 0.69)								

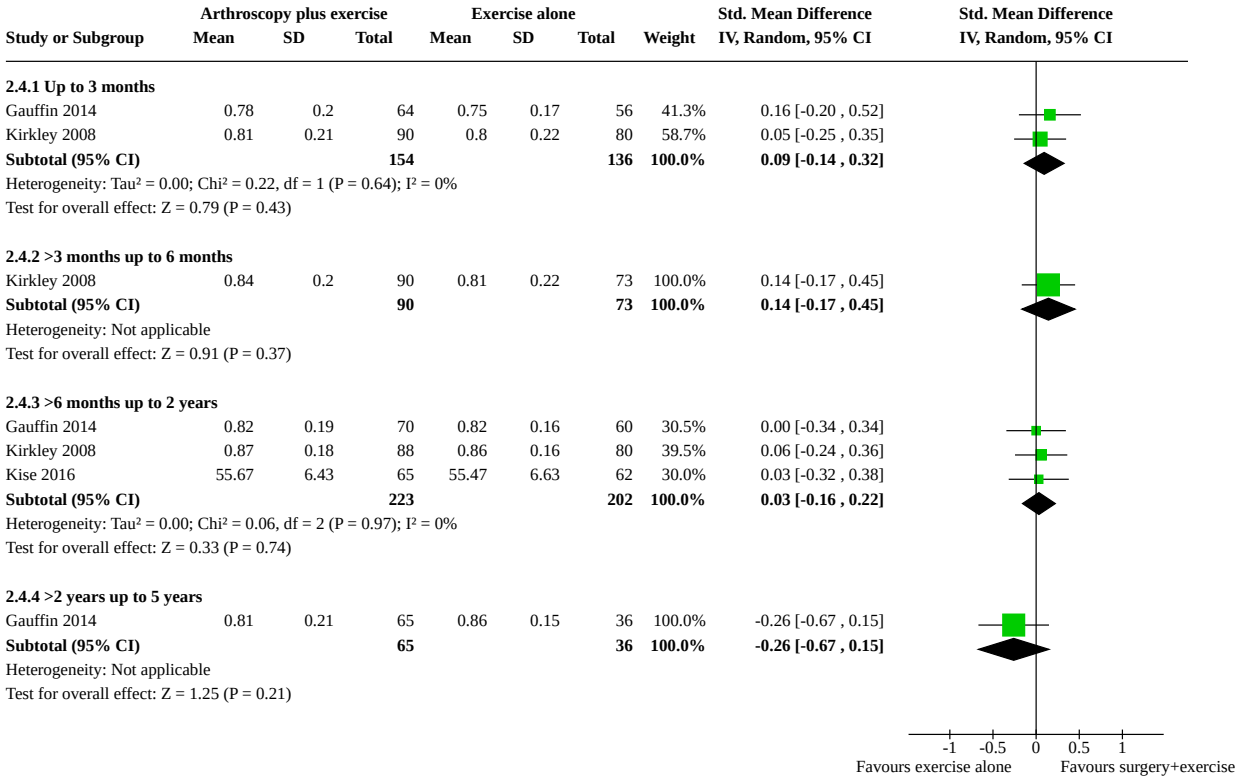


Analysis 2.3. Comparison 2: Arthroscopic surgery versus exercise, Outcome 3: Knee-specific quality of life (higher score=better)

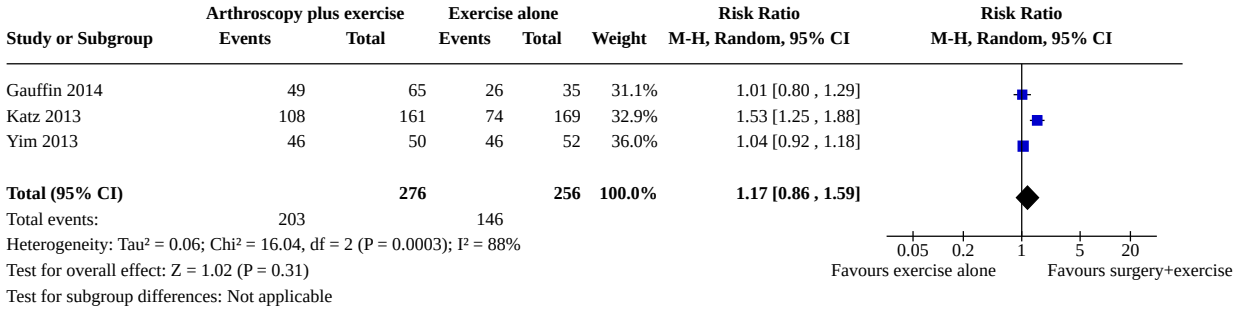
Study or Subgroup	Arthroscopy plus exercise Mean	SD	Total	Exercise alone Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
2.3.1 Up to 3 months								
Gauffin 2014	56	21.96	66	49	21.2	56	31.6%	7.00 [-0.67 , 14.67]
Herrlin 2007	60.83	20.09	47	56.92	17.33	49	33.0%	3.91 [-3.61 , 11.43]
Kise 2016	67.01	20.36	64	57.5	21.66	65	35.4%	9.51 [2.26 , 16.76]
Subtotal (95% CI)			**177**			**170**	**100.0%**	**6.87 [2.55 , 11.19]**
Heterogeneity: Tau² = 0.00; Chi² = 1.11, df = 2 (P = 0.58); I² = 0%								
Test for overall effect: Z = 3.12 (P = 0.002)								
2.3.2 >3 months up to 6 months								
Herrlin 2007	62.77	22.72	47	62.28	21.03	49	100.0%	0.49 [-8.28 , 9.26]
Subtotal (95% CI)			**47**			**49**	**100.0%**	**0.49 [-8.28 , 9.26]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.11 (P = 0.91)								
2.3.3 >6 months up to 2 years								
Gauffin 2014	66	26.89	70	59	24.7	60	30.3%	7.00 [-1.87 , 15.87]
Herrlin 2007	73.22	22.44	46	75.35	22.16	46	29.1%	-2.13 [-11.24 , 6.98]
Kise 2016	79.1	19.26	64	71.77	21.32	62	40.6%	7.33 [0.23 , 14.43]
Subtotal (95% CI)			**180**			**168**	**100.0%**	**4.47 [-1.33 , 10.28]**
Heterogeneity: Tau² = 8.50; Chi² = 2.94, df = 2 (P = 0.23); I² = 32%								
Test for overall effect: Z = 1.51 (P = 0.13)								
2.3.4 >2 years up to 5 years								
Gauffin 2014	65.3	24.26	65	68.1	26.48	36	39.9%	-2.80 [-13.27 , 7.67]
Kise 2016	78.9	21.08	62	73.5	20.99	57	60.1%	5.40 [-2.16 , 12.96]
Subtotal (95% CI)			**127**			**93**	**100.0%**	**2.13 [-5.74 , 10.00]**
Heterogeneity: Tau² = 11.91; Chi² = 1.55, df = 1 (P = 0.21); I² = 35%								
Test for overall effect: Z = 0.53 (P = 0.60)								



Analysis 2.4. Comparison 2: Arthroscopic surgery versus exercise, Outcome 4: Generic quality of life (higher score=better)



Analysis 2.5. Comparison 2: Arthroscopic surgery versus exercise, Outcome 5: Participant-reported success

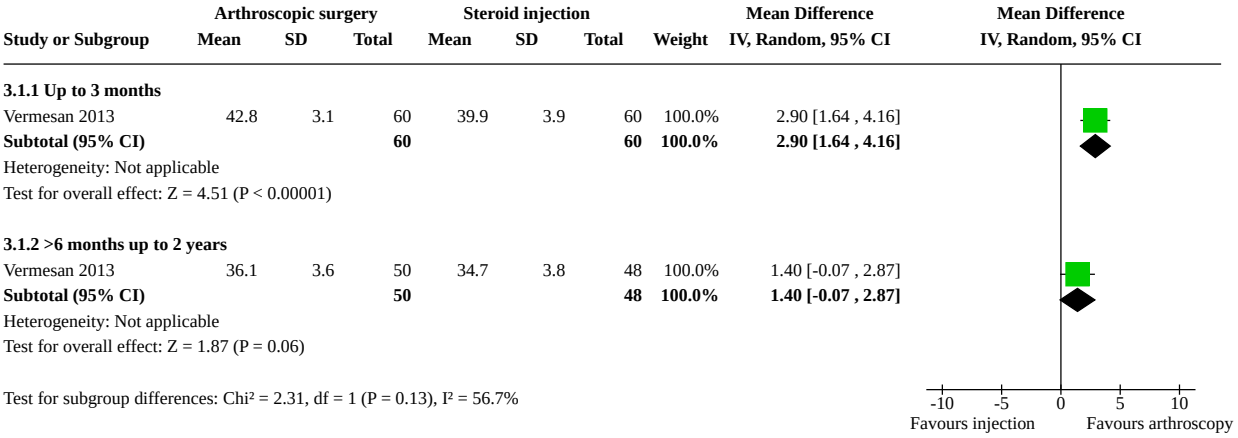


Comparison 3. Arthroscopic surgery versus glucocorticoid injection

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Function (OKS, 0-48, higher score=better function)	1		Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1.1 Up to 3 months	1	120	Mean Difference (IV, Random, 95% CI)	2.90 [1.64, 4.16]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1.2 >6 months up to 2 years	1	98	Mean Difference (IV, Random, 95% CI)	1.40 [-0.07, 2.87]

Analysis 3.1. Comparison 3: Arthroscopic surgery versus glucocorticoid injection, Outcome 1: Function (OKS, 0-48, higher score=better function)

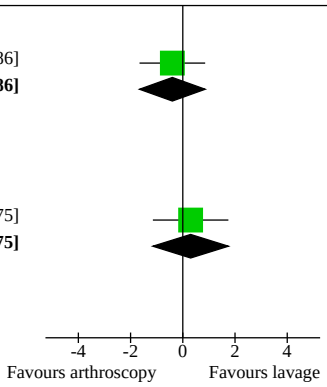


Comparison 4. Arthroscopic surgery versus non-arthroscopic lavage

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1 Pain (AIMS-P subscale, 0-10, lower score=less pain)	1		Mean Difference (IV, Random, 95% CI)	Subtotals only
4.1.1 Up to 3 months	1	32	Mean Difference (IV, Random, 95% CI)	-0.40 [-1.66, 0.86]
4.1.2 >6 months up to 2 years	1	32	Mean Difference (IV, Random, 95% CI)	0.30 [-1.15, 1.75]
4.2 Function (AIMS-PF subscale, 0-10, higher score=better function)	1		Mean Difference (IV, Random, 95% CI)	Subtotals only
4.2.1 Up to 3 months	1	32	Mean Difference (IV, Random, 95% CI)	0.50 [-0.25, 1.25]
4.2.2 >6 months up to 2 years	1	32	Mean Difference (IV, Random, 95% CI)	0.30 [-0.50, 1.10]
4.3 Participant-reported success (≥1cm improvement in VAS)	1		Risk Ratio (M-H, Random, 95% CI)	Subtotals only

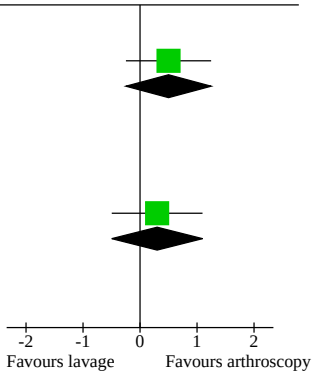
Analysis 4.1. Comparison 4: Arthroscopic surgery versus non-arthroscopic lavage, Outcome 1: Pain (AIMS-P subscale, 0-10, lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Non-arthroscopic lavage Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
4.1.1 Up to 3 months								
Chang 1993	5	1.8	18	5.4	1.8	14	100.0%	-0.40 [-1.66 , 0.86]
Subtotal (95% CI)			**18**			**14**	**100.0%**	**-0.40 [-1.66 , 0.86]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.62 (P = 0.53)								
4.1.2 >6 months up to 2 years								
Chang 1993	5.3	2.07	18	5	2.07	14	100.0%	0.30 [-1.15 , 1.75]
Subtotal (95% CI)			**18**			**14**	**100.0%**	**0.30 [-1.15 , 1.75]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.41 (P = 0.68)								
Test for subgroup differences: Chi² = 0.51, df = 1 (P = 0.47), I² = 0%								



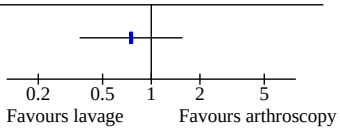
Analysis 4.2. Comparison 4: Arthroscopic surgery versus non-arthroscopic lavage, Outcome 2: Function (AIMS-PF subscale, 0-10, higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Non-arthroscopic lavage Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
4.2.1 Up to 3 months								
Chang 1993	-1.5	1.07	18	-2	1.07	14	100.0%	0.50 [-0.25 , 1.25]
Subtotal (95% CI)			**18**			**14**	**100.0%**	**0.50 [-0.25 , 1.25]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.31 (P = 0.19)								
4.2.2 >6 months up to 2 years								
Chang 1993	-1.7	1.14	18	-2	1.14	14	100.0%	0.30 [-0.50 , 1.10]
Subtotal (95% CI)			**18**			**14**	**100.0%**	**0.30 [-0.50 , 1.10]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.74 (P = 0.46)								
Test for subgroup differences: Chi² = 0.13, df = 1 (P = 0.72), I² = 0%								



Analysis 4.3. Comparison 4: Arthroscopic surgery versus non-arthroscopic lavage, Outcome 3: Participant-reported success (≥1cm improvement in VAS)

Study or Subgroup	Arthroscopic surgery Events	Total	Non-arthroscopic lavage Events	Total	Risk Ratio M-H, Random, 95% CI
Chang 1993	7	16	7	12	0.75 [0.36 , 1.56]



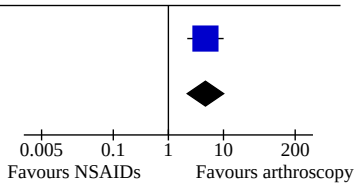
Comparison 5. Arthroscopic surgery versus NSAIDs

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.1 Participant-reported success	1	73	Risk Ratio (M-H, Random, 95% CI)	4.70 [2.20, 10.06]

Analysis 5.1. Comparison 5: Arthroscopic surgery versus NSAIDs, Outcome 1: Participant-reported success

Study or Subgroup	Arthroscopic surgery Events	Total	NSAIDs Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Merchan 1993	26	35	6	38	100.0%	4.70 [2.20 , 10.06]
Total (95% CI)		**35**		**38**	**100.0%**	**4.70 [2.20 , 10.06]**
Total events:	26		6			

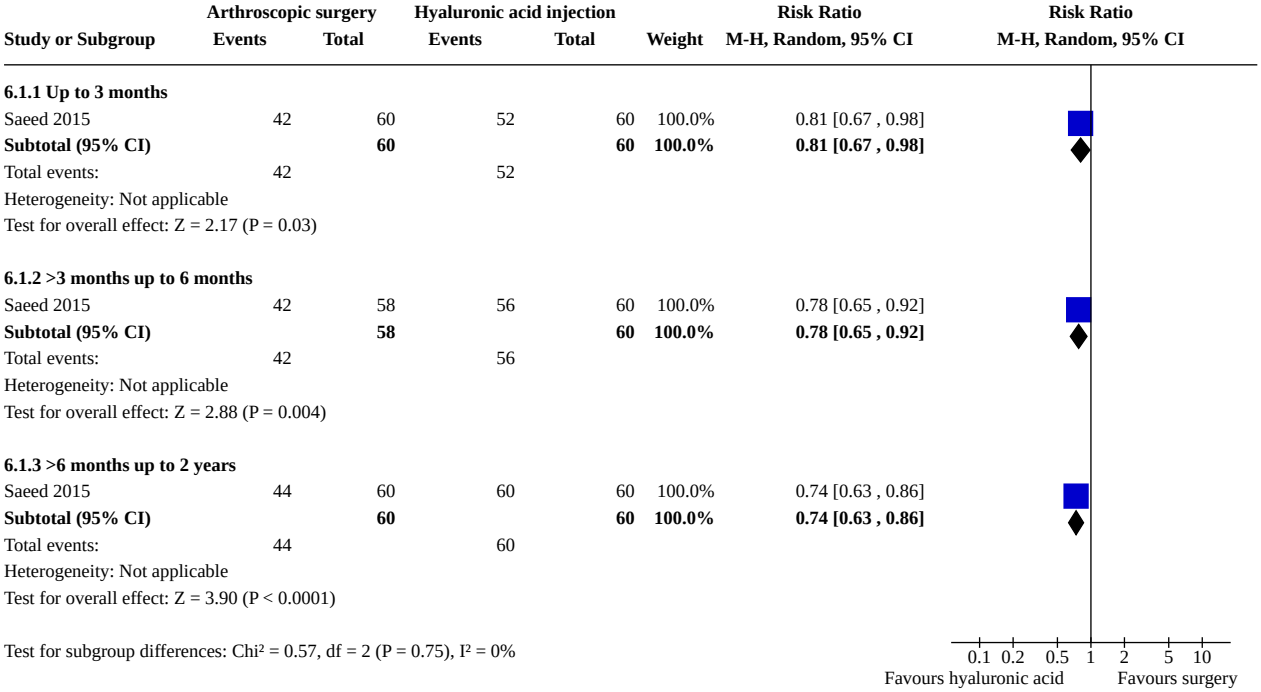
Heterogeneity: Not applicable
Test for overall effect: Z = 4.00 (P < 0.0001)
Test for subgroup differences: Not applicable



Comparison 6. Arthroscopic surgery versus hyaluronic acid injections

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6.1 Pain (KSSS pain score of 30 or higher; higher=less pain)	1		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
6.1.1 Up to 3 months	1	120	Risk Ratio (M-H, Random, 95% CI)	0.81 [0.67, 0.98]
6.1.2 >3 months up to 6 months	1	118	Risk Ratio (M-H, Random, 95% CI)	0.78 [0.65, 0.92]
6.1.3 >6 months up to 2 years	1	120	Risk Ratio (M-H, Random, 95% CI)	0.74 [0.63, 0.86]

Analysis 6.1. Comparison 6: Arthroscopic surgery versus hyaluronic acid injections, Outcome 1: Pain (KSSS pain score of 30 or higher; higher=less pain)



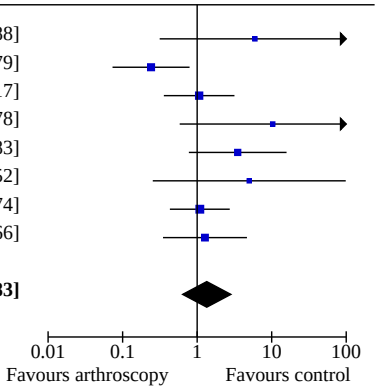
Comparison 7. Harms: arthroscopic surgery versus control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
7.1 Serious adverse events	8	1206	Risk Ratio (M-H, Random, 95% CI)	1.35 [0.64, 2.83]
7.2 Total adverse events	9	1326	Risk Ratio (M-H, Random, 95% CI)	1.15 [0.78, 1.70]
7.3 Progression of knee osteoarthritis	5	533	Risk Ratio (M-H, Random, 95% CI)	1.25 [1.01, 1.54]
7.4 Subsequent knee surgery (replacement or high tibial osteotomy)	4	864	Risk Ratio (M-H, Random, 95% CI)	2.63 [0.94, 7.34]

Analysis 7.1. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 1: Serious adverse events

Study or Subgroup	Arthroscopic surgery Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Gauffin 2014	3	66	0	56	5.3%	5.96 [0.31 , 112.88]
Herrlin 2007	3	47	13	49	16.8%	0.24 [0.07 , 0.79]
Katz 2013	8	164	5	109	18.1%	1.06 [0.36 , 3.17]
Kise 2016	5	64	0	60	5.5%	10.32 [0.58 , 182.78]
Merchan 1993	7	40	2	40	13.3%	3.50 [0.77 , 15.83]
Roos 2018	2	22	0	22	5.1%	5.00 [0.25 , 98.52]
Sihvonen 2013	8	70	8	76	20.3%	1.09 [0.43 , 2.74]
Van de Graaf 2018	5	159	4	162	15.6%	1.27 [0.35 , 4.66]
Total (95% CI)		**632**		**574**	**100.0%**	**1.35 [0.64 , 2.83]**
Total events:	41		32			

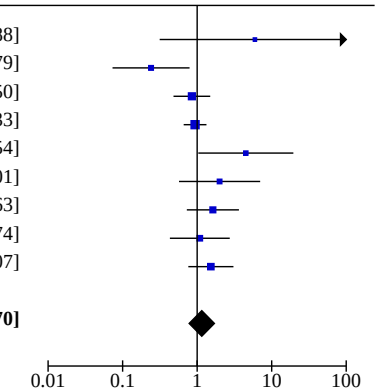
Heterogeneity: Tau² = 0.49; Chi² = 13.22, df = 7 (P = 0.07); I² = 47%
Test for overall effect: Z = 0.78 (P = 0.43)
Test for subgroup differences: Not applicable



Analysis 7.2. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 2: Total adverse events

Study or Subgroup	Arthroscopic surgery Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Gauffin 2014	3	66	0	56	1.6%	5.96 [0.31 , 112.88]
Herrlin 2007	3	47	13	49	7.7%	0.24 [0.07 , 0.79]
Katz 2013	23	164	18	109	17.4%	0.85 [0.48 , 1.50]
Kise 2016	31	64	31	60	22.4%	0.94 [0.66 , 1.33]
Merchan 1993	9	40	2	40	5.6%	4.50 [1.04 , 19.54]
Roos 2018	6	22	3	22	7.1%	2.00 [0.57 , 7.01]
Saeed 2015	13	60	8	60	12.7%	1.63 [0.73 , 3.63]
Sihvonen 2013	8	70	8	76	10.8%	1.09 [0.43 , 2.74]
Van de Graaf 2018	18	159	12	162	14.6%	1.53 [0.76 , 3.07]
Total (95% CI)		**692**		**634**	**100.0%**	**1.15 [0.78 , 1.70]**
Total events:	114		95			

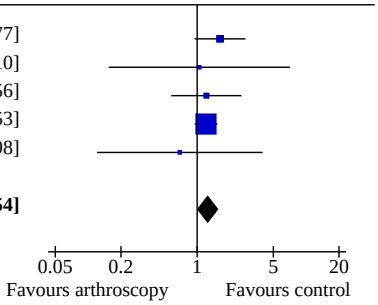
Heterogeneity: Tau² = 0.14; Chi² = 15.28, df = 8 (P = 0.05); I² = 48%
Test for overall effect: Z = 0.72 (P = 0.47)
Test for subgroup differences: Not applicable



Analysis 7.3. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 3: Progression of knee osteoarthritis

Study or Subgroup	Arthroscopic surgery Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Gauffin 2014	33	55	10	27	15.0%	1.62 [0.95 , 2.77]
Herrlin 2007	2	43	2	45	1.2%	1.05 [0.15 , 7.10]
Kise 2016	13	62	10	58	7.9%	1.22 [0.58 , 2.56]
Sihvonen 2013	48	67	44	74	74.5%	1.20 [0.95 , 1.53]
Yim 2013	2	50	3	52	1.4%	0.69 [0.12 , 3.98]
Total (95% CI)		**277**		**256**	**100.0%**	**1.25 [1.01 , 1.54]**
Total events:	98		69			

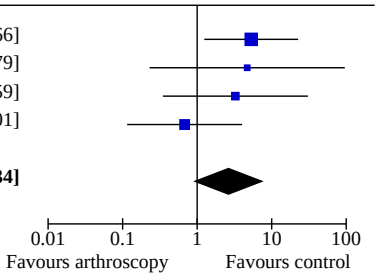
Heterogeneity: Tau² = 0.00; Chi² = 1.47, df = 4 (P = 0.83); I² = 0%
Test for overall effect: Z = 2.09 (P = 0.04)
Test for subgroup differences: Not applicable



Analysis 7.4. Comparison 7: Harms: arthroscopic surgery versus control, Outcome 4: Subsequent knee surgery (replacement or high tibial osteotomy)

Study or Subgroup	Arthroscopic surgery Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Katz 2013	16	164	2	109	40.8%	5.32 [1.25 , 22.66]
Kise 2016	2	64	0	60	11.0%	4.69 [0.23 , 95.79]
Sihvonen 2013	3	70	1	76	19.2%	3.26 [0.35 , 30.59]
Van de Graaf 2018	2	159	3	162	29.0%	0.68 [0.12 , 4.01]
Total (95% CI)		**457**		**407**	**100.0%**	**2.63 [0.94 , 7.34]**
Total events:	23		6			

Heterogeneity: Tau² = 0.13; Chi² = 3.37, df = 3 (P = 0.34); I² = 11%
Test for overall effect: Z = 1.84 (P = 0.07)
Test for subgroup differences: Not applicable



Comparison 8. Subgroup analysis: presence of meniscal tear

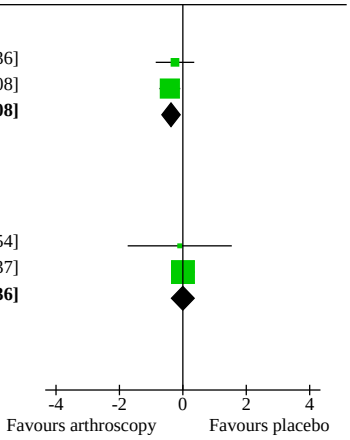
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
8.1 Pain up to 3 months (lower score=less pain)	4		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.1.1 Participants with meniscal tear	2	188	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-0.66, -0.08]
8.1.2 Unclear if participants had meniscal tear	2	121	Std. Mean Difference (IV, Random, 95% CI)	-0.00 [-0.36, 0.36]
8.2 Pain at >3 months up to 6 months (lower score=less pain)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.2.1 Participants with meniscal tear	1	146	Std. Mean Difference (IV, Random, 95% CI)	-0.25 [-0.58, 0.08]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
8.2.2 Unclear if participants had meniscal tear	2	119	Std. Mean Difference (IV, Random, 95% CI)	0.03 [-0.33, 0.39]
8.3 Pain at >6 months up to 2 years (lower score=less pain)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.3.1 Participants with meniscal tear	2	188	Std. Mean Difference (IV, Random, 95% CI)	-0.31 [-0.88, 0.25]
8.3.2 Unclear if participants had meniscal tear	1	107	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.49, 0.27]
8.4 Pain at >2 years up to 5 years (lower score=less pain)	1		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.4.1 Participants with meniscal tear	1	142	Std. Mean Difference (IV, Random, 95% CI)	-0.08 [-0.41, 0.24]
8.4.2 Unclear if participants had meniscal tear	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
8.5 Function up to 3 months (higher score=better function)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.5.1 Participants with meniscal tear	2	188	Std. Mean Difference (IV, Random, 95% CI)	0.08 [-0.20, 0.37]
8.5.2 Unclear if participants had meniscal tear	1	114	Std. Mean Difference (IV, Random, 95% CI)	-0.12 [-0.48, 0.25]
8.6 Function at >3 months up to 6 months (higher score=better function)	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.6.1 Participants with meniscal tear	1	145	Std. Mean Difference (IV, Random, 95% CI)	0.01 [-0.32, 0.33]
8.6.2 Unclear if participants had meniscal tear	1	112	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.27, 0.47]
8.7 Function at >6 months up to 2 years (higher score=better function)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.7.1 Participants with meniscal tear	2	187	Std. Mean Difference (IV, Random, 95% CI)	0.24 [-0.48, 0.96]
8.7.2 Unclear if participants had meniscal tear	1	106	Std. Mean Difference (IV, Random, 95% CI)	-0.04 [-0.42, 0.34]
8.8 Function at >2 years up to 5 years (higher score=better function)	1		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.8.1 Participants with meniscal tear	1	142	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.48, 0.18]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
8.8.2 Unclear if participants had meniscal tear	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable

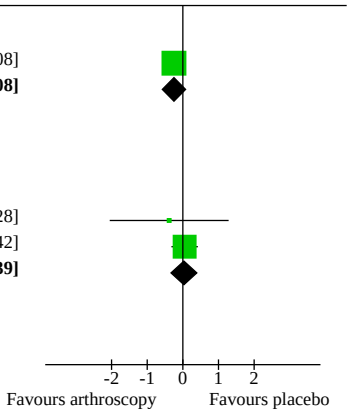
Analysis 8.1. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 1: Pain up to 3 months (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.1.1 Participants with meniscal tear								
Roos 2018	-71.9	17.6	21	-67	21	21	22.6%	-0.25 [-0.86 , 0.36]
Sihvonen 2013	3.1	2.13	70	4.1	2.67	76	77.4%	-0.41 [-0.74 , -0.08]
Subtotal (95% CI)			**91**			**97**	**100.0%**	**-0.37 [-0.66 , -0.08]**
Heterogeneity: Tau² = 0.00; Chi² = 0.21, df = 1 (P = 0.65); I² = 0%								
Test for overall effect: Z = 2.53 (P = 0.01)								
8.1.2 Unclear if participants had meniscal tear								
Moseley 1996	4.5	2.13	2	4.8	2.67	5	4.8%	-0.10 [-1.74 , 1.54]
Moseley 2002	-46.8	21.9	58	-46.9	24.9	56	95.2%	0.00 [-0.36 , 0.37]
Subtotal (95% CI)			**60**			**61**	**100.0%**	**-0.00 [-0.36 , 0.36]**
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.90); I² = 0%								
Test for overall effect: Z = 0.00 (P = 1.00)								
Test for subgroup differences: Chi² = 2.52, df = 1 (P = 0.11), I² = 60.3%								



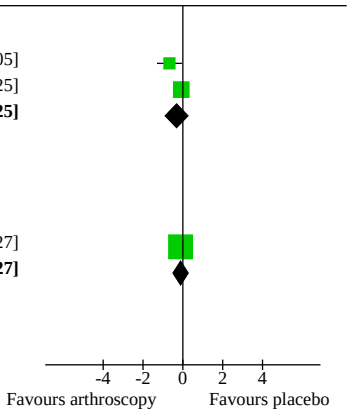
Analysis 8.2. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 2: Pain at >3 months up to 6 months (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.2.1 Participants with meniscal tear								
Sihvonen 2013	2.5	2.35	70	3.1	2.43	76	100.0%	-0.25 [-0.58 , 0.08]
Subtotal (95% CI)			**70**			**76**	**100.0%**	**-0.25 [-0.58 , 0.08]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.50 (P = 0.13)								
8.2.2 Unclear if participants had meniscal tear								
Moseley 1996	4.5	2.35	2	5.6	2.43	5	4.7%	-0.38 [-2.05 , 1.28]
Moseley 2002	-45.1	20.6	55	-46.3	26.4	57	95.3%	0.05 [-0.32 , 0.42]
Subtotal (95% CI)			**57**			**62**	**100.0%**	**0.03 [-0.33 , 0.39]**
Heterogeneity: Tau² = 0.00; Chi² = 0.25, df = 1 (P = 0.62); I² = 0%								
Test for overall effect: Z = 0.16 (P = 0.87)								



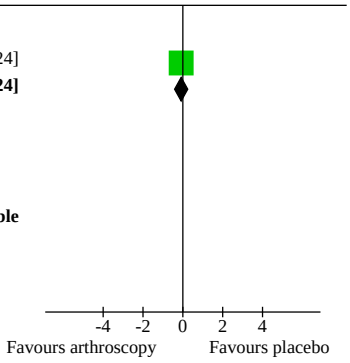
Analysis 8.3. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 3: Pain at >6 months up to 2 years (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.3.1 Participants with meniscal tear								
Roos 2018	-79.1	17.2	22	-63.1	28.6	20	39.5%	-0.67 [-1.30 , -0.05]
Sihvonen 2013	2.7	2.54	70	2.9	2.45	76	60.5%	-0.08 [-0.40 , 0.25]
Subtotal (95% CI)			**92**			**96**	**100.0%**	**-0.31 [-0.88 , 0.25]**
Heterogeneity: Tau² = 0.11; Chi² = 2.73, df = 1 (P = 0.10); I² = 63%								
Test for overall effect: Z = 1.08 (P = 0.28)								
8.3.2 Unclear if participants had meniscal tear								
Moseley 2002	-45	23	52	-42.3	24.2	55	100.0%	-0.11 [-0.49 , 0.27]
Subtotal (95% CI)			**52**			**55**	**100.0%**	**-0.11 [-0.49 , 0.27]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.59 (P = 0.56)								



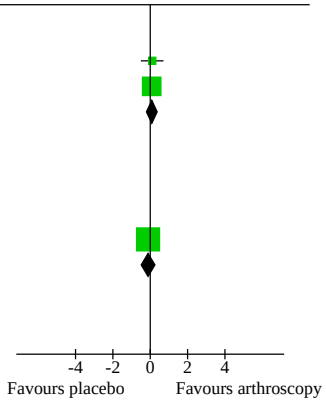
Analysis 8.4. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 4: Pain at >2 years up to 5 years (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.4.1 Participants with meniscal tear								
Sihvonen 2013	2	2.3	68	2.2	2.4	74	100.0%	-0.08 [-0.41 , 0.24]
Subtotal (95% CI)			**68**			**74**	**100.0%**	**-0.08 [-0.41 , 0.24]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.50 (P = 0.61)								
8.4.2 Unclear if participants had meniscal tear								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								



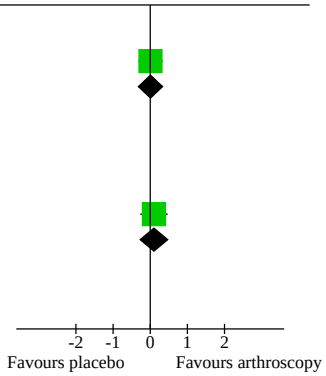
Analysis 8.5. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 5: Function up to 3 months (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.5.1 Participants with meniscal tear								
Roos 2018	76.7	15.7	21	74.8	19.8	21	22.4%	0.10 [-0.50 , 0.71]
Sihvonen 2013	77.2	15.79	70	75.9	17.35	76	77.6%	0.08 [-0.25 , 0.40]
Subtotal (95% CI)			**91**			**97**	**100.0%**	**0.08 [-0.20 , 0.37]**
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.94); I² = 0%								
Test for overall effect: Z = 0.57 (P = 0.57)								
8.5.2 Unclear if participants had meniscal tear								
Moseley 2002	49.6	24.2	58	52.4	23.5	56	100.0%	-0.12 [-0.48 , 0.25]
Subtotal (95% CI)			**58**			**56**	**100.0%**	**-0.12 [-0.48 , 0.25]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.62 (P = 0.53)								
Test for subgroup differences: Chi² = 0.71, df = 1 (P = 0.40), I² = 0%								



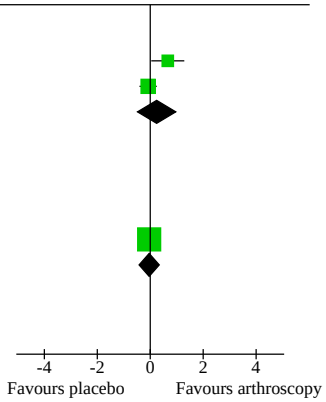
Analysis 8.6. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 6: Function at >3 months up to 6 months (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.6.1 Participants with meniscal tear								
Sihvonen 2013	82.8	16.01	70	82.7	14.58	75	100.0%	0.01 [-0.32 , 0.33]
Subtotal (95% CI)			**70**			**75**	**100.0%**	**0.01 [-0.32 , 0.33]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.04 (P = 0.97)								
8.6.2 Unclear if participants had meniscal tear								
Moseley 2002	51	25.9	55	48.4	25.9	57	100.0%	0.10 [-0.27 , 0.47]
Subtotal (95% CI)			**55**			**57**	**100.0%**	**0.10 [-0.27 , 0.47]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.53 (P = 0.60)								
Test for subgroup differences: Chi² = 0.14, df = 1 (P = 0.71), I² = 0%								



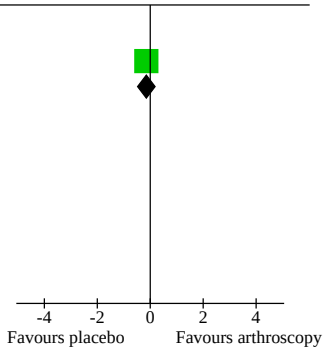
Analysis 8.7. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 7: Function at >6 months up to 2 years (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.7.1 Participants with meniscal tear								
Roos 2018	85.1	15.5	22	71.2	25	20	43.3%	0.66 [0.04 , 1.29]
Sihvonen 2013	82.2	15.89	69	83.4	13.79	76	56.7%	-0.08 [-0.41 , 0.25]
Subtotal (95% CI)			**91**			**96**	**100.0%**	**0.24 [-0.48 , 0.96]**
Heterogeneity: Tau² = 0.21; Chi² = 4.29, df = 1 (P = 0.04); I² = 77%								
Test for overall effect: Z = 0.66 (P = 0.51)								
8.7.2 Unclear if participants had meniscal tear								
Moseley 2002	47.9	26.6	52	49	27.2	54	100.0%	-0.04 [-0.42 , 0.34]
Subtotal (95% CI)			**52**			**54**	**100.0%**	**-0.04 [-0.42 , 0.34]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.21 (P = 0.83)								
Test for subgroup differences: Chi² = 0.46, df = 1 (P = 0.50), I² = 0%								



Analysis 8.8. Comparison 8: Subgroup analysis: presence of meniscal tear, Outcome 8: Function at >2 years up to 5 years (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.8.1 Participants with meniscal tear								
Sihvonen 2013	83.7	14.3	68	85.8	14	74	100.0%	-0.15 [-0.48 , 0.18]
Subtotal (95% CI)			**68**			**74**	**100.0%**	**-0.15 [-0.48 , 0.18]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.88 (P = 0.38)								
8.8.2 Unclear if participants had meniscal tear								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
Test for subgroup differences: Not applicable								



Comparison 9. Subgroup analysis: arthroscopy with supervised exercise

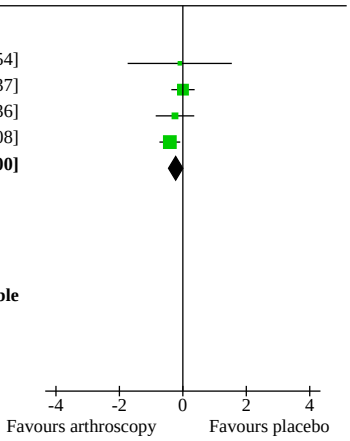
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
9.1 Pain up to 3 months (lower score=less pain)	4		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
9.1.1 Unsupervised/ home exercises	4	309	Std. Mean Difference (IV, Random, 95% CI)	-0.23 [-0.45, -0.00]
9.1.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
9.2 Pain at >3 months up to 6 months (lower score=less pain)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
9.2.1 Unsupervised/ home exercises	3	265	Std. Mean Difference (IV, Random, 95% CI)	-0.12 [-0.37, 0.12]
9.2.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
9.3 Pain at >6 months up to 2 years (lower score=less pain)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
9.3.1 Unsupervised/ home exercises	3	295	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.48, 0.09]
9.3.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
9.4 Pain at >2 years up to 5 years (lower score=less pain)	1		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
9.4.1 Unsupervised/ home exercises	1	142	Std. Mean Difference (IV, Random, 95% CI)	-0.08 [-0.41, 0.24]
9.4.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
9.5 Function up to 3 months (higher score=better function)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
9.5.1 Unsupervised/ home exercises	3	302	Std. Mean Difference (IV, Random, 95% CI)	0.01 [-0.22, 0.23]
9.5.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
9.6 Function at >3 months up to 6 months (higher score=better function)	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
9.6.1 Unsupervised/ home exercises	2	257	Std. Mean Difference (IV, Random, 95% CI)	0.05 [-0.20, 0.29]
9.6.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
9.7 Function at >6 months up to 2 years (higher score=better function)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
9.7.1 Unsupervised/ home exercises	3	293	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.27, 0.47]
9.7.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
9.8 Function at >2 years up to 5 years (higher score=better function)	1		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
9.8.1 Unsupervised/ home exercises	1	142	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.48, 0.18]
9.8.2 Supervised exercise	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable

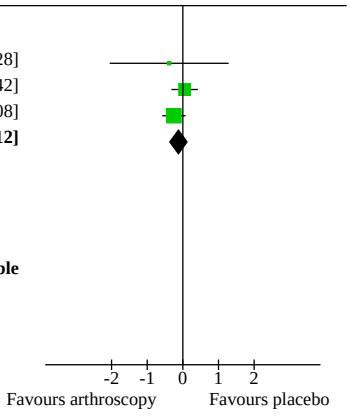
Analysis 9.1. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 1: Pain up to 3 months (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.1.1 Unsupervised/ home exercises								
Moseley 1996	4.5	2.13	2	4.8	2.67	5	1.9%	-0.10 [-1.74 , 1.54]
Moseley 2002	-46.8	21.9	58	-46.9	24.9	56	37.5%	0.00 [-0.36 , 0.37]
Roos 2018	-71.9	17.6	21	-67	21	21	13.7%	-0.25 [-0.86 , 0.36]
Sihvonen 2013	3.1	2.13	70	4.1	2.67	76	46.9%	-0.41 [-0.74 , -0.08]
Subtotal (95% CI)			**151**			**158**	**100.0%**	**-0.23 [-0.45 , -0.00]**
Heterogeneity: Tau² = 0.00; Chi² = 2.75, df = 3 (P = 0.43); I² = 0%								
Test for overall effect: Z = 1.98 (P = 0.05)								
9.1.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								



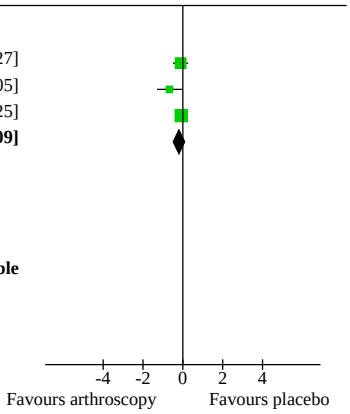
Analysis 9.2. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 2: Pain at >3 months up to 6 months (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.2.1 Unsupervised/ home exercises								
Moseley 1996	4.5	2.35	2	5.6	2.43	5	2.1%	-0.38 [-2.05 , 1.28]
Moseley 2002	-45.1	20.6	55	-46.3	26.4	57	42.7%	0.05 [-0.32 , 0.42]
Sihvonen 2013	2.5	2.35	70	3.1	2.43	76	55.2%	-0.25 [-0.58 , 0.08]
Subtotal (95% CI)			**127**			**138**	**100.0%**	**-0.12 [-0.37 , 0.12]**
Heterogeneity: Tau² = 0.00; Chi² = 1.51, df = 2 (P = 0.47); I² = 0%								
Test for overall effect: Z = 1.01 (P = 0.31)								
9.2.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								



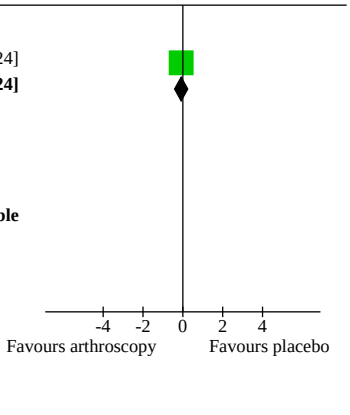
Analysis 9.3. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 3: Pain at >6 months up to 2 years (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.3.1 Unsupervised/ home exercises								
Moseley 2002	-45	23	52	-42.3	24.2	55	37.3%	-0.11 [-0.49 , 0.27]
Roos 2018	-79.1	17.2	22	-63.1	28.6	20	17.7%	-0.67 [-1.30 , -0.05]
Sihvonen 2013	2.7	2.54	70	2.9	2.45	76	45.1%	-0.08 [-0.40 , 0.25]
Subtotal (95% CI)			**144**			**151**	**100.0%**	**-0.20 [-0.48 , 0.09]**
Heterogeneity: Tau² = 0.02; Chi² = 2.88, df = 2 (P = 0.24); I² = 30%								
Test for overall effect: Z = 1.35 (P = 0.18)								
9.3.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								



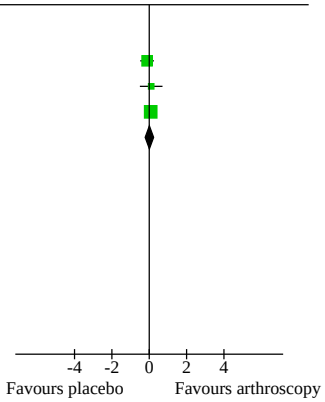
Analysis 9.4. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 4: Pain at >2 years up to 5 years (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.4.1 Unsupervised/ home exercises								
Sihvonen 2013	2	2.3	68	2.2	2.4	74	100.0%	-0.08 [-0.41 , 0.24]
Subtotal (95% CI)			**68**			**74**	**100.0%**	**-0.08 [-0.41 , 0.24]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.50 (P = 0.61)								
9.4.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								



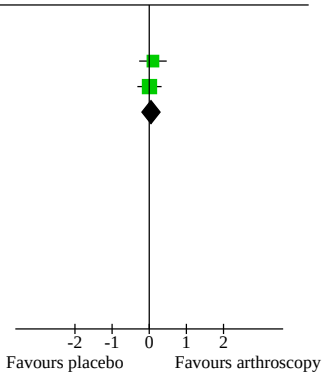
Analysis 9.5. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 5: Function up to 3 months (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.5.1 Unsupervised/ home exercises								
Moseley 2002	49.6	24.2	58	52.4	23.5	56	37.8%	-0.12 [-0.48 , 0.25]
Roos 2018	76.7	15.7	21	74.8	19.8	21	13.9%	0.10 [-0.50 , 0.71]
Sihvonen 2013	77.2	15.79	70	75.9	17.35	76	48.3%	0.08 [-0.25 , 0.40]
Subtotal (95% CI)			**149**			**153**	**100.0%**	**0.01 [-0.22 , 0.23]**
Heterogeneity: Tau² = 0.00; Chi² = 0.72, df = 2 (P = 0.70); I² = 0%								
Test for overall effect: Z = 0.07 (P = 0.94)								
9.5.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
Test for subgroup differences: Not applicable								



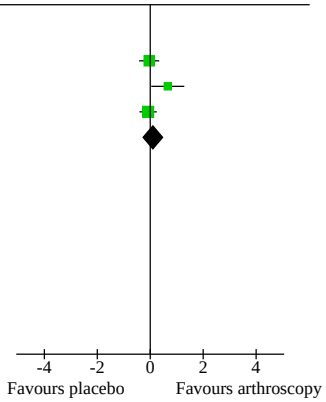
Analysis 9.6. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 6: Function at >3 months up to 6 months (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.6.1 Unsupervised/ home exercises								
Moseley 2002	51	25.9	55	48.4	25.9	57	43.6%	0.10 [-0.27 , 0.47]
Sihvonen 2013	82.8	16.01	70	82.7	14.58	75	56.4%	0.01 [-0.32 , 0.33]
Subtotal (95% CI)			**125**			**132**	**100.0%**	**0.05 [-0.20 , 0.29]**
Heterogeneity: Tau² = 0.00; Chi² = 0.14, df = 1 (P = 0.71); I² = 0%								
Test for overall effect: Z = 0.38 (P = 0.71)								
9.6.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
Test for subgroup differences: Not applicable								



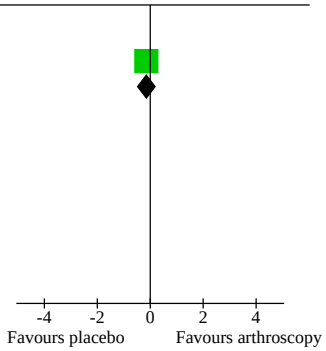
Analysis 9.7. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 7: Function at >6 months up to 2 years (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.7.1 Unsupervised/ home exercises								
Moseley 2002	47.9	26.6	52	49	27.2	54	36.8%	-0.04 [-0.42 , 0.34]
Roos 2018	85.1	15.5	22	71.2	25	20	22.1%	0.66 [0.04 , 1.29]
Sihvonen 2013	82.2	15.89	69	83.4	13.79	76	41.1%	-0.08 [-0.41 , 0.25]
Subtotal (95% CI)			**143**			**150**	**100.0%**	**0.10 [-0.27 , 0.47]**
Heterogeneity: Tau² = 0.06; Chi² = 4.53, df = 2 (P = 0.10); I² = 56%								
Test for overall effect: Z = 0.53 (P = 0.60)								
9.7.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
Test for subgroup differences: Not applicable								



Analysis 9.8. Comparison 9: Subgroup analysis: arthroscopy with supervised exercise, Outcome 8: Function at >2 years up to 5 years (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
9.8.1 Unsupervised/ home exercises								
Sihvonen 2013	83.7	14.3	68	85.8	14	74	100.0%	-0.15 [-0.48 , 0.18]
Subtotal (95% CI)			**68**			**74**	**100.0%**	**-0.15 [-0.48 , 0.18]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.88 (P = 0.38)								
9.8.2 Supervised exercise								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
Test for subgroup differences: Not applicable								



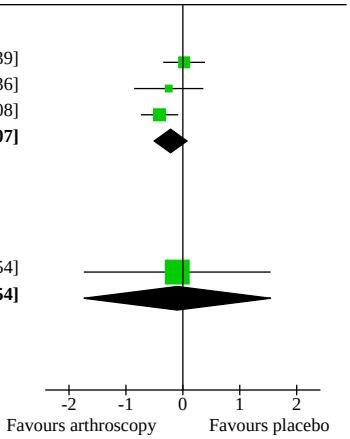
Comparison 10. Sensitivity analysis: low risk of selection bias

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
10.1 Pain up to 3 months (lower score=less pain)	4		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
10.1.1 Low risk of bias	3	302	Std. Mean Difference (IV, Random, 95% CI)	-0.22 [-0.50, 0.07]
10.1.2 Risk of bias	1	7	Std. Mean Difference (IV, Random, 95% CI)	-0.10 [-1.74, 1.54]
10.2 Function up to 3 months (higher score=better function)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
10.2.1 Low risk of bias	3	302	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.24, 0.46]

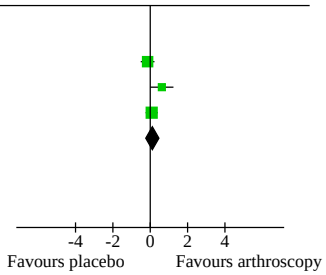
Analysis 10.1. Comparison 10: Sensitivity analysis: low risk of selection bias, Outcome 1: Pain up to 3 months (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
10.1.1 Low risk of bias								
Moseley 2002	49.3	22	58	48.8	21.5	56	38.1%	0.02 [-0.34 , 0.39]
Roos 2018	-71.9	17.6	21	-67	21	21	18.4%	-0.25 [-0.86 , 0.36]
Sihvonen 2013	3.1	2.13	70	4.1	2.67	76	43.5%	-0.41 [-0.74 , -0.08]
Subtotal (95% CI)			**149**			**153**	**100.0%**	**-0.22 [-0.50 , 0.07]**
Heterogeneity: Tau² = 0.02; Chi² = 2.98, df = 2 (P = 0.23); I² = 33%								
Test for overall effect: Z = 1.47 (P = 0.14)								
10.1.2 Risk of bias								
Moseley 1996	4.5	2.13	2	4.8	2.67	5	100.0%	-0.10 [-1.74 , 1.54]
Subtotal (95% CI)			**2**			**5**	**100.0%**	**-0.10 [-1.74 , 1.54]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.12 (P = 0.91)								



Analysis 10.2. Comparison 10: Sensitivity analysis: low risk of selection bias, Outcome 2: Function up to 3 months (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
10.2.1 Low risk of bias								
Moseley 2002	-53.5	28.6	58	-49.9	21.6	56	37.5%	-0.14 [-0.51 , 0.23]
Roos 2018	72.5	15.7	21	61	20.3	21	21.4%	0.62 [0.00 , 1.24]
Sihvonen 2013	77.2	15.79	70	75.9	17.35	76	41.2%	0.08 [-0.25 , 0.40]
Subtotal (95% CI)			**149**			**153**	**100.0%**	**0.11 [-0.24 , 0.46]**
Heterogeneity: Tau² = 0.05; Chi² = 4.30, df = 2 (P = 0.12); I² = 53%								
Test for overall effect: Z = 0.62 (P = 0.53)								
Test for subgroup differences: Not applicable								



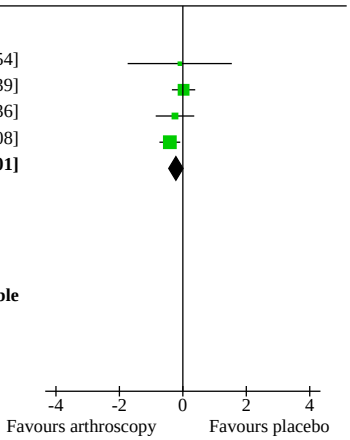
Comparison 11. Sensitivity analysis: low risk of detection bias

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
11.1 Pain up to 3 months (lower score=less pain)	4		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
11.1.1 Low risk of bias	4	309	Std. Mean Difference (IV, Random, 95% CI)	-0.22 [-0.44, 0.01]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
11.1.2 Risk of bias	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
11.2 Function up to 3 months (higher score=better function)	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
11.2.1 Low risk	3	302	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.24, 0.46]
11.2.2 Risk of bias	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable

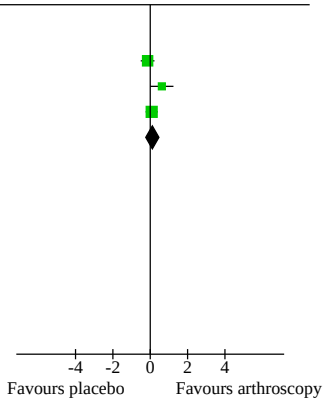
Analysis 11.1. Comparison 11: Sensitivity analysis: low risk of detection bias, Outcome 1: Pain up to 3 months (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
11.1.1 Low risk of bias								
Moseley 1996	4.5	2.13	2	4.8	2.67	5	1.9%	-0.10 [-1.74 , 1.54]
Moseley 2002	49.3	22	58	48.8	21.5	56	37.5%	0.02 [-0.34 , 0.39]
Roos 2018	-71.9	17.6	21	-67	21	21	13.7%	-0.25 [-0.86 , 0.36]
Sihvonen 2013	3.1	2.13	70	4.1	2.67	76	46.9%	-0.41 [-0.74 , -0.08]
Subtotal (95% CI)			**151**			**158**	**100.0%**	**-0.22 [-0.44 , 0.01]**
Heterogeneity: Tau² = 0.00; Chi² = 3.00, df = 3 (P = 0.39); I² = 0%								
Test for overall effect: Z = 1.92 (P = 0.06)								
11.1.2 Risk of bias								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								



Analysis 11.2. Comparison 11: Sensitivity analysis: low risk of detection bias, Outcome 2: Function up to 3 months (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
11.2.1 Low risk								
Moseley 2002	-53.5	28.6	58	-49.9	21.6	56	37.5%	-0.14 [-0.51 , 0.23]
Roos 2018	72.5	15.7	21	61	20.3	21	21.4%	0.62 [0.00 , 1.24]
Sihvonen 2013	77.2	15.79	70	75.9	17.35	76	41.2%	0.08 [-0.25 , 0.40]
Subtotal (95% CI)			**149**			**153**	**100.0%**	**0.11 [-0.24 , 0.46]**
Heterogeneity: Tau² = 0.05; Chi² = 4.30, df = 2 (P = 0.12); I² = 53%								
Test for overall effect: Z = 0.62 (P = 0.53)								
11.2.2 Risk of bias								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
Test for subgroup differences: Not applicable								

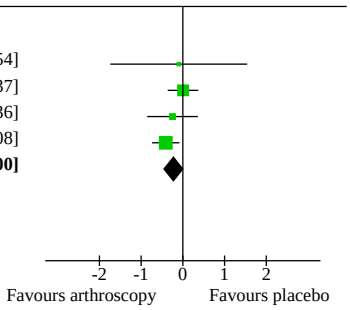


Comparison 12. Sensitivity analysis: fixed-effect model

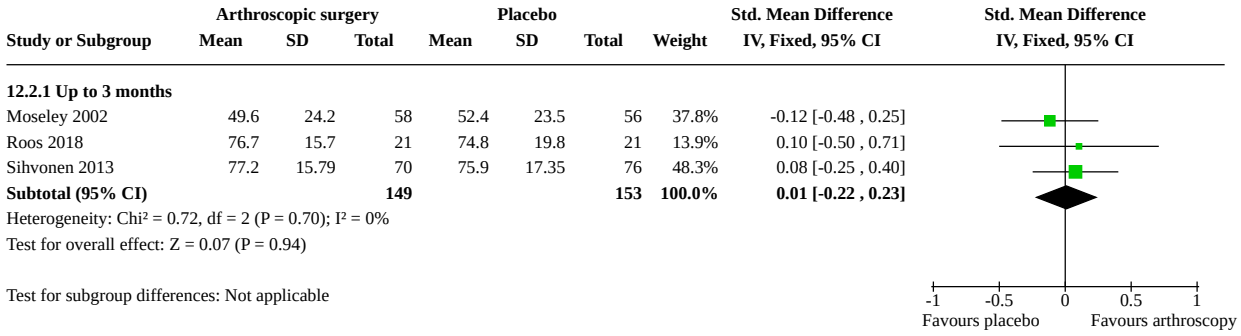
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
12.1 Pain (lower score=less pain)	4		Std. Mean Difference (IV, Fixed, 95% CI)	Subtotals only
12.1.1 Up to 3 months	4	309	Std. Mean Difference (IV, Fixed, 95% CI)	-0.23 [-0.45, -0.00]
12.2 Function (higher score=better function)	3		Std. Mean Difference (IV, Fixed, 95% CI)	Subtotals only
12.2.1 Up to 3 months	3	302	Std. Mean Difference (IV, Fixed, 95% CI)	0.01 [-0.22, 0.23]

Analysis 12.1. Comparison 12: Sensitivity analysis: fixed-effect model, Outcome 1: Pain (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
12.1.1 Up to 3 months								
Moseley 1996	4.5	2.13	2	4.8	2.67	5	1.9%	-0.10 [-1.74 , 1.54]
Moseley 2002	-46.8	21.9	58	-46.9	24.9	56	37.5%	0.00 [-0.36 , 0.37]
Roos 2018	-71.9	17.6	21	-67	21	21	13.7%	-0.25 [-0.86 , 0.36]
Sihvonen 2013	3.1	2.13	70	4.1	2.67	76	46.9%	-0.41 [-0.74 , -0.08]
Subtotal (95% CI)			**151**			**158**	**100.0%**	**-0.23 [-0.45 , -0.00]**
Heterogeneity: Chi² = 2.75, df = 3 (P = 0.43); I² = 0%								
Test for overall effect: Z = 1.98 (P = 0.05)								



Analysis 12.2. Comparison 12: Sensitivity analysis: fixed-effect model, Outcome 2: Function (higher score=better function)



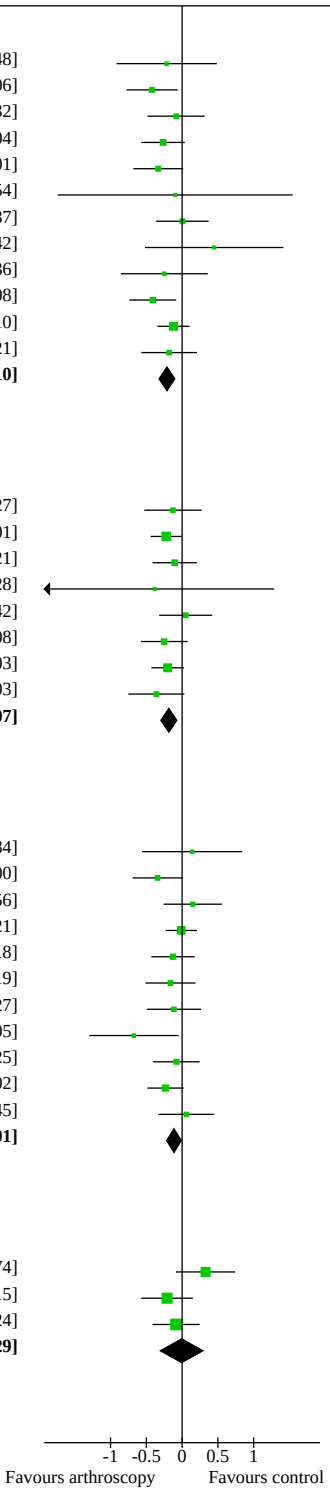
Comparison 13. Sensitivity analysis: arthroscopic surgery versus any control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
13.1 Pain (lower score=less pain)	13		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
13.1.1 Up to 3 months	12	1283	Std. Mean Difference (IV, Random, 95% CI)	-0.21 [-0.32, -0.10]
13.1.2 >3 months up to 6 months	8	1252	Std. Mean Difference (IV, Random, 95% CI)	-0.19 [-0.30, -0.07]
13.1.3 >6 months up to 2 years	11	1505	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.22, -0.01]
13.1.4 >2 years up to 5 years	3	361	Std. Mean Difference (IV, Random, 95% CI)	-0.01 [-0.30, 0.29]
13.2 Function (higher score=better function)	13		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
13.2.1 Up to 3 months	12	1403	Std. Mean Difference (IV, Random, 95% CI)	0.19 [0.04, 0.34]
13.2.2 >3 months up to 6 months	7	1245	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.01, 0.21]
13.2.3 >6 months up to 2 years	12	1651	Std. Mean Difference (IV, Random, 95% CI)	0.11 [0.01, 0.20]
13.2.4 >2 years up to 5 years	3	361	Std. Mean Difference (IV, Random, 95% CI)	-0.09 [-0.30, 0.12]
13.3 Knee-specific quality of life (higher score=better)	5		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
13.3.1 Up to 3 months	5	535	Std. Mean Difference (IV, Random, 95% CI)	0.33 [0.16, 0.50]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
13.3.2 >3 months up to 6 months	2	242	Std. Mean Difference (IV, Random, 95% CI)	0.08 [-0.17, 0.33]
13.3.3 >6 months up to 2 years	5	536	Std. Mean Difference (IV, Random, 95% CI)	0.19 [-0.00, 0.38]
13.3.4 >2 years up to 5 years	3	362	Std. Mean Difference (IV, Random, 95% CI)	0.05 [-0.16, 0.26]
13.3.5 >5 years up to 10 years	0	0	Std. Mean Difference (IV, Random, 95% CI)	Not estimable
13.4 Generic quality of life (higher score=better)	5		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
13.4.1 Up to 3 months	3	332	Std. Mean Difference (IV, Random, 95% CI)	-0.03 [-0.36, 0.30]
13.4.2 >3 months up to 6 months	1	163	Std. Mean Difference (IV, Random, 95% CI)	0.14 [-0.17, 0.45]
13.4.3 >6 months up to 2 years	5	613	Std. Mean Difference (IV, Random, 95% CI)	0.08 [-0.08, 0.24]
13.4.4 >2 years up to 5 years	1	101	Std. Mean Difference (IV, Random, 95% CI)	-0.26 [-0.67, 0.15]
13.5 Participant-reported success	8	851	Risk Ratio (M-H, Random, 95% CI)	1.24 [0.96, 1.60]

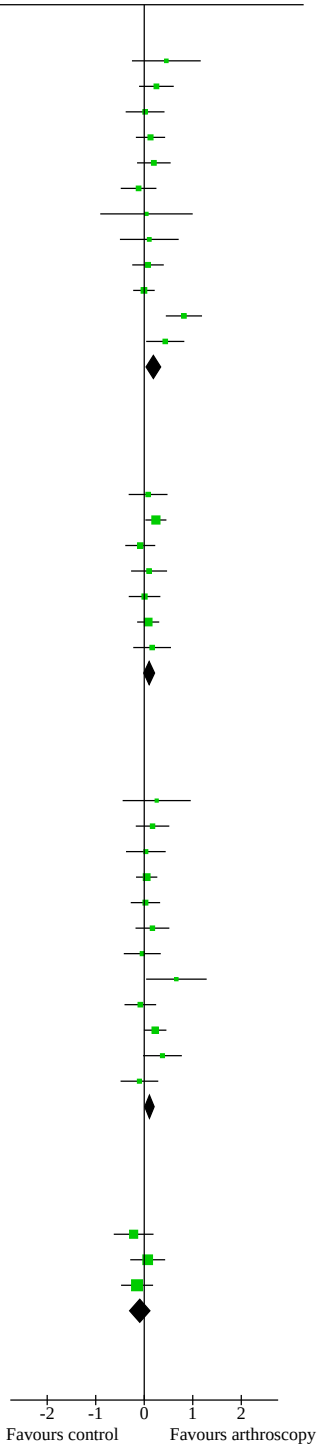
Analysis 13.1. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 1: Pain (lower score=less pain)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
13.1.1 Up to 3 months								
Chang 1993	5	1.8	18	5.4	1.8	14	2.5%	-0.22 [-0.92 , 0.48]
Gauffin 2014	-77	16.6	66	-69	21.2	57	9.4%	-0.42 [-0.78 , -0.06]
Herrlin 2007	-81.9	16.2	47	-80.6	14.8	49	7.6%	-0.08 [-0.48 , 0.32]
Kirkley 2008	141	109	90	172	124	80	13.2%	-0.27 [-0.57 , 0.04]
Kise 2016	-81.94	16.87	64	-75.64	20.53	65	10.0%	-0.33 [-0.68 , 0.01]
Moseley 1996	4.5	2.13	2	4.8	2.67	5	0.4%	-0.10 [-1.74 , 1.54]
Moseley 2002	-46.8	21.9	58	-46.9	24.9	56	9.0%	0.00 [-0.36 , 0.37]
Osteras 2012	2.6	1.1	8	2	1.4	9	1.3%	0.45 [-0.52 , 1.42]
Roos 2018	-71.9	17.6	21	-67	21	21	3.3%	-0.25 [-0.86 , 0.36]
Sihvonen 2013	3.1	2.13	70	4.1	2.67	76	11.3%	-0.41 [-0.74 , -0.08]
Van de Graaf 2018	30.4	24.6	154	33.4	25.4	151	24.0%	-0.12 [-0.34 , 0.10]
Yim 2013	2.4	1.8	50	2.7	1.5	52	8.0%	-0.18 [-0.57 , 0.21]
Subtotal (95% CI)			**648**			**635**	**100.0%**	**-0.21 [-0.32 , -0.10]**
Heterogeneity: Tau² = 0.00; Chi² = 7.52, df = 11 (P = 0.76); I² = 0%								
Test for overall effect: Z = 3.77 (P = 0.0002)								
13.1.2 >3 months up to 6 months								
Herrlin 2007	-83.7	15.7	47	-81.6	16.9	49	7.7%	-0.13 [-0.53 , 0.27]
Katz 2013	21.1	18.1	161	25.2	18.6	169	26.4%	-0.22 [-0.44 , -0.01]
Kirkley 2008	143	113	90	155	118	73	13.0%	-0.10 [-0.41 , 0.21]
Moseley 1996	4.5	2.35	2	5.6	2.43	5	0.4%	-0.38 [-2.05 , 1.28]
Moseley 2002	-45.1	20.6	55	-46.3	26.4	57	9.0%	0.05 [-0.32 , 0.42]
Sihvonen 2013	2.5	2.35	70	3.1	2.43	76	11.6%	-0.25 [-0.58 , 0.08]
Van de Graaf 2018	25.4	27.7	151	31	27.7	145	23.7%	-0.20 [-0.43 , 0.03]
Yim 2013	1.5	1.8	50	2.1	1.5	52	8.1%	-0.36 [-0.75 , 0.03]
Subtotal (95% CI)			**626**			**626**	**100.0%**	**-0.19 [-0.30 , -0.07]**
Heterogeneity: Tau² = 0.00; Chi² = 3.00, df = 7 (P = 0.88); I² = 0%								
Test for overall effect: Z = 3.26 (P = 0.001)								
13.1.3 >6 months up to 2 years								
Chang 1993	5.3	2.07	18	5	2.07	14	2.1%	0.14 [-0.56 , 0.84]
Gauffin 2014	-84	14.9	70	-78	19.8	60	8.5%	-0.34 [-0.69 , 0.00]
Herrlin 2007	-87.8	16.4	46	-90.1	14	46	6.1%	0.15 [-0.26 , 0.56]
Katz 2013	19.1	17.5	156	19.3	17.6	164	21.4%	-0.01 [-0.23 , 0.21]
Kirkley 2008	168	134	88	185	132	80	11.2%	-0.13 [-0.43 , 0.18]
Kise 2016	-89.37	16	64	-86.69	16.84	62	8.4%	-0.16 [-0.51 , 0.19]
Moseley 2002	-45	23	52	-42.3	24.2	55	7.1%	-0.11 [-0.49 , 0.27]
Roos 2018	-79.1	17.2	22	-63.1	28.6	20	2.6%	-0.67 [-1.30 , -0.05]
Sihvonen 2013	2.7	2.54	70	2.9	2.45	76	9.7%	-0.08 [-0.40 , 0.25]
Van de Graaf 2018	19.6	23.8	115	25.5	26.9	125	15.9%	-0.23 [-0.49 , 0.02]
Yim 2013	1.8	1.8	50	1.7	1.5	52	6.8%	0.06 [-0.33 , 0.45]
Subtotal (95% CI)			**751**			**754**	**100.0%**	**-0.11 [-0.22 , -0.01]**
Heterogeneity: Tau² = 0.00; Chi² = 9.42, df = 10 (P = 0.49); I² = 0%								
Test for overall effect: Z = 2.20 (P = 0.03)								
13.1.4 >2 years up to 5 years								
Gauffin 2014	-79.3	20.97	65	-86	18.88	35	29.2%	0.33 [-0.09 , 0.74]
Kise 2016	-90	14.86	62	-86.7	16.37	57	33.8%	-0.21 [-0.57 , 0.15]
Sihvonen 2013	2	2.3	68	2.2	2.4	74	36.9%	-0.08 [-0.41 , 0.24]
Subtotal (95% CI)			**195**			**166**	**100.0%**	**-0.01 [-0.30 , 0.29]**
Heterogeneity: Tau² = 0.03; Chi² = 3.94, df = 2 (P = 0.14); I² = 49%								
Test for overall effect: Z = 0.04 (P = 0.97)								
Test for subgroup differences: Chi² = 2.87, df = 3 (P = 0.41), I² = 0%								



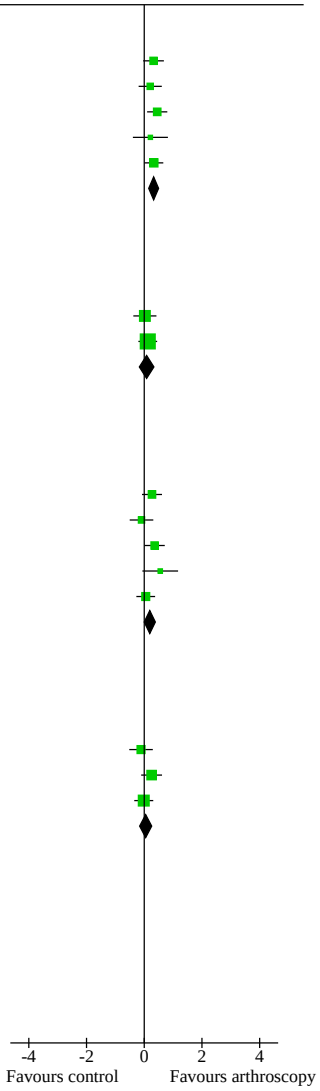
Analysis 13.2. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 2: Function (higher score=better function)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
13.2.1 Up to 3 months								
Chang 1993	-1.5	1.07	18	-2	1.07	14	3.7%	0.46 [-0.25 , 1.16]
Gauffin 2014	81	18.5	65	76	21.2	57	9.4%	0.25 [-0.11 , 0.61]
Herrlin 2007	86.51	16.85	47	86.22	15.93	49	8.3%	0.02 [-0.38 , 0.42]
Kirkley 2008	-522	341	90	-568	369	80	11.1%	0.13 [-0.17 , 0.43]
Kise 2016	88.72	16.18	64	85.25	18.55	65	9.7%	0.20 [-0.15 , 0.54]
Moseley 2002	49.6	24.2	58	52.4	23.5	56	9.1%	-0.12 [-0.48 , 0.25]
Osteras 2012	40.9	23.1	8	39.7	25.9	9	2.2%	0.05 [-0.91 , 1.00]
Roos 2018	76.7	15.7	21	74.8	19.8	21	4.7%	0.10 [-0.50 , 0.71]
Sihvonen 2013	77.2	15.79	70	75.9	17.35	76	10.4%	0.08 [-0.25 , 0.40]
Van de Graaf 2018	59.9	16.6	155	60	17.4	158	14.0%	-0.01 [-0.23 , 0.22]
Vermesan 2013	42.8	3.1	60	39.9	3.9	60	9.0%	0.82 [0.44 , 1.19]
Yim 2013	85.2	11.2	50	80.4	10.8	52	8.5%	0.43 [0.04 , 0.83]
Subtotal (95% CI)			**706**			**697**	**100.0%**	**0.19 [0.04 , 0.34]**
Heterogeneity: Tau² = 0.03; Chi² = 19.96, df = 11 (P = 0.05); I² = 45%								
Test for overall effect: Z = 2.45 (P = 0.01)								
13.2.2 >3 months up to 6 months								
Herrlin 2007	88.4	16.13	47	87.06	17.33	49	7.7%	0.08 [-0.32 , 0.48]
Katz 2013	-14.7	17.8	161	-19	17.9	169	26.4%	0.24 [0.02 , 0.46]
Kirkley 2008	-551	382	90	-520	368	73	13.0%	-0.08 [-0.39 , 0.23]
Moseley 2002	51	25.9	55	48.4	25.9	57	9.0%	0.10 [-0.27 , 0.47]
Sihvonen 2013	82.8	16.01	70	82.7	14.58	75	11.7%	0.01 [-0.32 , 0.33]
Van de Graaf 2018	64.7	19.2	151	63.2	17.5	146	23.9%	0.08 [-0.15 , 0.31]
Yim 2013	84.1	11.2	50	82.3	10.8	52	8.2%	0.16 [-0.23 , 0.55]
Subtotal (95% CI)			**624**			**621**	**100.0%**	**0.10 [-0.01 , 0.21]**
Heterogeneity: Tau² = 0.00; Chi² = 3.40, df = 6 (P = 0.76); I² = 0%								
Test for overall effect: Z = 1.79 (P = 0.07)								
13.2.3 >6 months up to 2 years								
Chang 1993	-1.7	1.14	18	-2	1.14	14	1.9%	0.26 [-0.45 , 0.96]
Gauffin 2014	86	17.1	70	83	17.8	60	7.8%	0.17 [-0.17 , 0.52]
Herrlin 2007	91.63	14.73	46	91.13	15.02	46	5.6%	0.03 [-0.38 , 0.44]
Katz 2013	-13.7	15.9	156	-14.5	16	164	19.5%	0.05 [-0.17 , 0.27]
Kirkley 2008	-612	448	88	-623	439	80	10.2%	0.02 [-0.28 , 0.33]
Kise 2016	93.53	15.72	65	90.94	14.75	62	7.7%	0.17 [-0.18 , 0.52]
Moseley 2002	47.9	26.6	52	49	27.2	54	6.5%	-0.04 [-0.42 , 0.34]
Roos 2018	85.1	15.5	22	71.2	25	20	2.4%	0.66 [0.04 , 1.29]
Sihvonen 2013	82.2	15.89	69	83.4	13.79	76	8.8%	-0.08 [-0.41 , 0.25]
Van de Graaf 2018	71.5	16.4	141	67.7	17.2	148	17.5%	0.23 [-0.01 , 0.46]
Vermesan 2013	36.1	3.6	50	34.7	3.8	48	5.9%	0.38 [-0.02 , 0.78]
Yim 2013	83.2	11.2	50	84.3	10.8	52	6.2%	-0.10 [-0.49 , 0.29]
Subtotal (95% CI)			**827**			**824**	**100.0%**	**0.11 [0.01 , 0.20]**
Heterogeneity: Tau² = 0.00; Chi² = 9.82, df = 11 (P = 0.55); I² = 0%								
Test for overall effect: Z = 2.17 (P = 0.03)								
13.2.4 >2 years up to 5 years								
Gauffin 2014	83.9	18.78	64	87.9	16.99	36	26.1%	-0.22 [-0.63 , 0.19]
Kise 2016	93	14.66	62	92	13.1	57	33.8%	0.07 [-0.29 , 0.43]
Sihvonen 2013	83.7	14.3	68	85.8	14	74	40.2%	-0.15 [-0.48 , 0.18]
Subtotal (95% CI)			**194**			**167**	**100.0%**	**-0.09 [-0.30 , 0.12]**
Heterogeneity: Tau² = 0.00; Chi² = 1.27, df = 2 (P = 0.53); I² = 0%								
Test for overall effect: Z = 0.86 (P = 0.39)								
Test for subgroup differences: Chi² = 4.58, df = 3 (P = 0.20), I² = 34.6%								



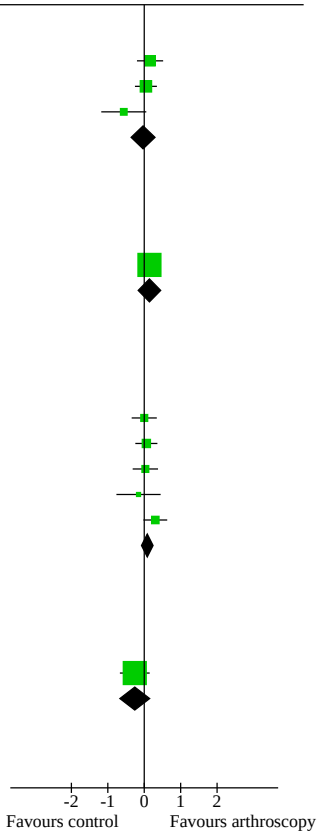
Analysis 13.3. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 3: Knee-specific quality of life (higher score=better)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
13.3.1 Up to 3 months								
Gauffin 2014	56	21.96	66	49	21.2	56	22.7%	0.32 [-0.04 , 0.68]
Herrlin 2007	60.83	20.09	47	56.92	17.33	49	18.1%	0.21 [-0.19 , 0.61]
Kise 2016	67.01	20.36	64	57.5	21.66	65	23.9%	0.45 [0.10 , 0.80]
Roos 2018	50.1	16.7	21	46	20.8	21	7.9%	0.21 [-0.39 , 0.82]
Sihvonen 2013	76.8	17.93	70	69.7	23.8	76	27.3%	0.33 [0.01 , 0.66]
Subtotal (95% CI)			**268**			**267**	**100.0%**	**0.33 [0.16 , 0.50]**
Heterogeneity: Tau² = 0.00; Chi² = 0.95, df = 4 (P = 0.92); I² = 0%								
Test for overall effect: Z = 3.74 (P = 0.0002)								
13.3.2 >3 months up to 6 months								
Herrlin 2007	62.77	22.72	47	62.28	21.03	49	39.7%	0.02 [-0.38 , 0.42]
Sihvonen 2013	81.7	21.13	70	79.1	21.21	76	60.3%	0.12 [-0.20 , 0.45]
Subtotal (95% CI)			**117**			**125**	**100.0%**	**0.08 [-0.17 , 0.33]**
Heterogeneity: Tau² = 0.00; Chi² = 0.14, df = 1 (P = 0.70); I² = 0%								
Test for overall effect: Z = 0.64 (P = 0.52)								
13.3.3 >6 months up to 2 years								
Gauffin 2014	66	26.89	70	59	24.7	60	23.7%	0.27 [-0.08 , 0.62]
Herrlin 2007	73.22	22.44	46	75.35	22.16	46	18.2%	-0.09 [-0.50 , 0.31]
Kise 2016	79.1	19.26	64	71.77	21.32	62	23.1%	0.36 [0.01 , 0.71]
Roos 2018	63.5	21.1	22	49.4	28.6	20	8.9%	0.55 [-0.06 , 1.17]
Sihvonen 2013	81	20.77	70	79.9	21.35	76	26.1%	0.05 [-0.27 , 0.38]
Subtotal (95% CI)			**272**			**264**	**100.0%**	**0.19 [-0.00 , 0.38]**
Heterogeneity: Tau² = 0.01; Chi² = 4.97, df = 4 (P = 0.29); I² = 20%								
Test for overall effect: Z = 1.95 (P = 0.05)								
13.3.4 >2 years up to 5 years								
Gauffin 2014	65.3	24.26	65	68.1	26.48	36	26.3%	-0.11 [-0.52 , 0.30]
Kise 2016	78.9	21.08	62	73.5	20.99	57	33.5%	0.26 [-0.11 , 0.62]
Sihvonen 2013	84.3	17.8	68	84.6	17.9	74	40.3%	-0.02 [-0.35 , 0.31]
Subtotal (95% CI)			**195**			**167**	**100.0%**	**0.05 [-0.16 , 0.26]**
Heterogeneity: Tau² = 0.00; Chi² = 1.99, df = 2 (P = 0.37); I² = 0%								
Test for overall effect: Z = 0.46 (P = 0.64)								
13.3.5 >5 years up to 10 years								
Subtotal (95% CI)			**0**			**0**		**Not estimable**
Heterogeneity: Not applicable								
Test for overall effect: Not applicable								
Test for subgroup differences: Chi² = 4.87, df = 3 (P = 0.18), I² = 38.4%								



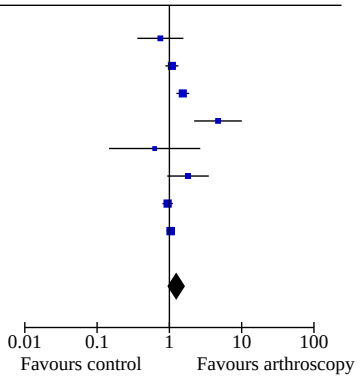
Analysis 13.4. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 4: Generic quality of life (higher score=better)

Study or Subgroup	Arthroscopic surgery Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
13.4.1 Up to 3 months								
Gauffin 2014	0.78	0.2	64	0.75	0.17	56	37.2%	0.16 [-0.20 , 0.52]
Kirkley 2008	0.81	0.21	90	0.8	0.22	80	42.9%	0.05 [-0.25 , 0.35]
Roos 2018	57.8	6.5	21	61.3	5.7	21	19.9%	-0.56 [-1.18 , 0.06]
Subtotal (95% CI)			**175**			**157**	**100.0%**	**-0.03 [-0.36 , 0.30]**
Heterogeneity: Tau² = 0.04; Chi² = 4.01, df = 2 (P = 0.13); I² = 50%								
Test for overall effect: Z = 0.19 (P = 0.85)								
13.4.2 >3 months up to 6 months								
Kirkley 2008	0.84	0.2	90	0.81	0.22	73	100.0%	0.14 [-0.17 , 0.45]
Subtotal (95% CI)			**90**			**73**	**100.0%**	**0.14 [-0.17 , 0.45]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.91 (P = 0.37)								
13.4.3 >6 months up to 2 years								
Gauffin 2014	0.82	0.19	70	0.82	0.16	60	21.2%	0.00 [-0.34 , 0.34]
Kirkley 2008	0.87	0.18	88	0.86	0.16	80	27.5%	0.06 [-0.24 , 0.36]
Kise 2016	55.67	6.43	65	55.47	6.63	62	20.8%	0.03 [-0.32 , 0.38]
Roos 2018	55.7	6.9	22	57	9.3	20	6.9%	-0.16 [-0.76 , 0.45]
Sihvonen 2013	0.94	0.06	70	0.92	0.07	76	23.6%	0.30 [-0.02 , 0.63]
Subtotal (95% CI)			**315**			**298**	**100.0%**	**0.08 [-0.08 , 0.24]**
Heterogeneity: Tau² = 0.00; Chi² = 2.70, df = 4 (P = 0.61); I² = 0%								
Test for overall effect: Z = 1.03 (P = 0.30)								
13.4.4 >2 years up to 5 years								
Gauffin 2014	0.81	0.21	65	0.86	0.15	36	100.0%	-0.26 [-0.67 , 0.15]
Subtotal (95% CI)			**65**			**36**	**100.0%**	**-0.26 [-0.67 , 0.15]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.25 (P = 0.21)								
Test for subgroup differences: Chi² = 2.96, df = 3 (P = 0.40), I² = 0%								



Analysis 13.5. Comparison 13: Sensitivity analysis: arthroscopic surgery versus any control, Outcome 5: Participant-reported success

Study or Subgroup	Arthroscopic surgery Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Chang 1993	7	16	7	12	7.6%	0.75 [0.36 , 1.56]
Gauffin 2014	54	70	42	59	17.9%	1.08 [0.88 , 1.33]
Katz 2013	108	161	74	169	18.0%	1.53 [1.25 , 1.88]
Merchan 1993	26	35	6	38	7.2%	4.70 [2.20 , 10.06]
Moseley 1996	1	2	4	5	2.7%	0.63 [0.15 , 2.67]
Roos 2018	14	21	7	19	8.5%	1.81 [0.93 , 3.51]
Sihvonen 2013	53	68	61	74	18.7%	0.95 [0.80 , 1.11]
Yim 2013	46	50	46	52	19.3%	1.04 [0.92 , 1.18]
Total (95% CI)		**423**		**428**	**100.0%**	**1.24 [0.96 , 1.60]**
Total events:	309		247			
Heterogeneity: Tau² = 0.08; Chi² = 42.33, df = 7 (P < 0.00001); I² = 83%						
Test for overall effect: Z = 1.67 (P = 0.09)						
Test for subgroup differences: Not applicable						



ADDITIONAL TABLES

Table 1. Characteristics of participants

Study ID	Age range (years)	Osteoarthritis and criteria	Meniscal tear and criteria
Trials with a placebo control			
Moseley 1996	< 70	ACR-defined or clinically diagnosed	No criteria re presence/absence specified
Moseley 2002	< 70	ACR-defined or clinically diagnosed	No criteria re presence/absence specified
Sihvonen 2013	35 to 65	KL grade 0 to 1	Medial meniscal tear on MRI
Roos 2018	35 to 55	KL grade 0 to 2	Medial meniscal tear on MRI
Trials with an exercise control			
Gauffin 2014	45 to 64	< 50% joint narrowing	Clinically suspected meniscal injury (66/75 randomised to surgery received surgery but only 56 had partial meniscectomies (1 loose bodies removed, 1 synovectomy, 1 partial resection ACL remnant, 8 deemed surgery unnecessary; of the 16/75 who crossed over to surgery, 11 had 11 partial meniscectomies (1 loose bodies removed, 1 microfracture, 1 partial resection ACL remnant, 1 deemed surgery unnecessary and 2 unknown). 3 ACL total ruptures were found (2 surgical group, 1 in non-surgery group who crossed over toe surgery)
Herrlin 2007	45 to 65	Grade 0 to 1 Ahlbacks classification	Medial meniscal tear
Katz 2013	45 or older meniscal tear, mild to mod OA, KL grade 0-3	KL grade 0 to 3	Meniscal tear
Kirkley 2008	18 or older	KL grade 2 to 4 except grade 4 if involves both compartments	Exclude large meniscal tear (bucket handle tear) mainly clinical, few had MRIs
Kise 2016	35 to 60	KL grade 0 to 2	Medial meniscal tear
Osteras 2012	35 to 60	KL grade 0 to 2	Meniscal tear
Van de Graaf 2018	45 to 70	KL grade 0 to 3	Meniscal tear
Yim 2013	no age restriction specified	KL grade 0 to 1	Medial posterior horn horizontal meniscal tear on MRI
Trials with other controls			
Saeed 2015	>40	KL grade 2 and 3	No criteria re presence/absence specified
Vermesan 2013	Not specified	medial compartment cartilage and meniscus lesions on MRI	Medial compartment cartilage and meniscus lesions on MRI

Table 1. Characteristics of participants *(Continued)*

Chang 1993	>20	KL grades 1 to 3	No criteria re presence/absence specified
Merchan 1993	Not specified	minimal joint space narrowing and formation of small osteophytes	No criteria re presence/absence specified

ACR: American College of Rheumatology; **ACL**: anterior cruciate ligament; **KL grade:** Kellgren-Lawrence classification grade; **MRI**: magnetic resonance imaging

Table 2. Characteristics of interventions used in included trials

Study ID	Description of arthroscopic surgery	Description of post-surgical exercise (in arthroscopic surgery arm)	Description of control	Co-interventions
Trials with a placebo control				
Moseley 1996	Arthroscopic debridement: diagnostic arthroscopy, joint lavage, shaving of rough articular cartilage, removal of loose debris, trimming of torn/degenerated menisci	Participants were instructed to resume walking and other activities of daily living as soon as their symptoms would allow. No other exercises were given.	Skin incisions without insertion of arthroscope, knee manipulation, saline splashing over the joint. Surgeon asked for all instruments. Simulation of standard arthroscopic debridement as close as possible. Time spent in the operation theatre: 1 hour	Oral analgesia (acetaminophen with codeine), crutches until able to walk comfortably without a limp. NSAIDs taken pre-operatively could be resumed after the first follow-up at 10 days. *Applied equally in all treatment groups:* yes
Moseley 2002	Arthroscopic debridement: diagnostic arthroscopy, joint lavage, shaving of rough articular cartilage, removal of loose debris, trimming of torn or degenerated meniscal fragments, and smoothening of the remaining meniscus to a firm and stable rim. Shaving of spurs from the tibial spine area that blocked full extension.	Participants were given a graduated exercise program after surgery; details of the program were not reported.	Simulated debridement with three 1 cm skin incisions but without insertion of the arthroscope. Knee manipulation, surgeon asked for all instruments, saline splashing. Time spent in the operation theatre: same as debridement group.	Walking aids, graduated exercise program and analgesics. *Applied equally in all treatment groups*: yes

Table 2. Characteristics of interventions used in included trials *(Continued)*

Roos 2018	Partial meniscectomy with preservation of as much meniscus as possible. Documentation of findings in cartilage, ligaments, synovium and the medial and lateral menisci. Registration of the type and extent of meniscus lesion and ICRS classification of articular cartilage changes.	Post-operative home-based exercise program. At 1 week, biking, swimming and fast walking, and at 2 to 3 weeks, more intense biking and jogging were recommended. For the first post-operative week, 7 different non-weight-bearing exercises to improve lower extremity function and knee range of motion were suggested, and an additional 3 weight-bearing thereafter. All exercises were recommended to be performed 10 to 15 times three times daily.	Skin incisions in same location as in arthroscopic surgery without insertion of arthroscope, knee manipulation, spillage of water, use of all equipment needed for arthroscopic surgery. Surgeon asked for all instruments. Simulation of arthroscopic surgery as close as possible.	Weight-bearing and non-weight-bearing exercises. *Applied equally in all treatment groups:* yes
Sihvonen 2013	Arthroscopic partial meniscectomy - removal of damaged menisci with arthroscopic instruments (mechanised shaver and meniscal punches) until solid meniscus tissue was reached. Resection of loose, unstable meniscal fragments while preserving as much of the meniscus tissue as possible.	Post-operative graduated home exercise program for both legs for 10 to 15 minutes at a time, 5 days a week.	The surgeon asked for all instruments, knee manipulation, simulation of a standard arthroscopic partial meniscectomy procedure by using a mechanised shaver (without the blade) outside the knee, suction was also used to drain the joint and saline was splashed. Time spent in the operation theatre: same as the surgery group.	Walking aids, graduated home exercise program, over-the-counter analgesics. *Applied equally in all treatment groups:* yes
Trials with an exercise control				
Gauffin 2014	Arthroscopic surgery: inspection of joint, meniscal resection performed if needed (but not performed if not needed)	Post-operatively all participants were allowed immediate, full weight-bearing activity. They were advised to resume the exercise programme according to	Unsupervised exercise program lasted 3 months, performed twice a week and comprised two phases. **Phase 1** was performed for the first 3 weeks and included 20 to 30 min brisk walk, 10 x 2 sets of the following: squats, pelvic lifts, pelvic lifts with ball between knees, heel raise, wall squats and standing on a pillow on one leg; **Phase 2:** 20 to 30 min brisk walk, 10 x 3 sets of all exercises done in phase 1.	None specified

Table 2. Characteristics of interventions used in included trials *(Continued)*

		phase 1 for 1 week, and then switch to phase 2.	*Frequency, intensity and duration:* phase 1 - daily, 2 sets; phase 2 - twice per week 3 sets each for 3 months. *Supervised:* no *Setting:* home	
Herrlin 2007	Arthroscopic partial meniscectomy: arthroscopic joint inspection, registration of meniscal lesions and Outerbridge classification of changes in the articular cartilage.	Twice a week during a period of 8 weeks each participant followed a standardised exercise program similar to the exercise group. This was followed by a written unsupervised home program twice a week.	**A. Supervised exercise.** *Description:* all exercises for 3 x 10 sets. 0 to 8 weeks: stationary bicycling 7 to 15 min, knee extensions concentrically with two legs and eccentrically with one leg, stair walking and balance on wobble boards (3 min), jogging, jumps, landing on a rebounder (5 min), stretching of knee extensors and flexors (1 min/muscle group). 0 to 4 weeks: calf raise on leg press, knee flexions concentrically with two legs and eccentrically with one leg. 1 to 4 weeks: leg press. 5 to 8 weeks: calf raises standing on one leg, lunges with < 80 of knee flexion with or without weight in the hands, knee flexions with one leg, knee extensions with one leg. *Frequency, intensity and duration:* twice a week for 8 weeks *Supervised*: yes. *Setting:* research centre. **B. Unsupervised exercise.** *Description*: 3 x 10 sets of one-leg standing during 1 min and a step down exercise. *Frequency, intensity and duration:* twice a week for 8 weeks. *Supervised:* no. *Setting:* home	None specified
Katz 2013	Arthroscopic partial meniscectomy: trimming of damaged meniscus to a stable rim, removal of loose fragments of cartilage and bone without any penetration of the subchondral bone	Post-operative standardised physical therapy program, as described in the exercise group.	**Supervised exercise.** *Description:* **phase I:** acute phase (1 to 10 days post-op) Retrograde Massage, Cryotherapy E-Stim: NMES or IFC, Joint Mobilisation Soft Tissue Mobilisation Stretching LE Muscles, Quad Sets SAQ/LAQ/HS Curls Hip-4 way, Bicycle, Elliptical, Treadmill, Leg Press, Balance/Proprioception. **Phase II:** Subacute Phase (10 days to 4 weeks post-op) Retrograde Massage Cryotherapy E-Stim: NMES or IFC, Joint Mobilisation Soft Tissue Mobilisation Stretching LE Muscles, Concentric/Eccentric Hip/Knee progressive resistive exercises, ROM, Resisted terminal knee extension, modified mini squats, step up/down progressions, toe raises, functional and agility training. **Phase III:** Advanced Activity Phase (4 to 7 weeks post-op) - continued stretching program, continued PRE therapeutic exercises program, closed chain program with progression to dynamic single leg stance, plyometrics, running, and sport specificity training. *Frequency, intensity and duration:* 8 exercises, 12 to 15 repetitions, 1 to 2 sets. *Supervised:* yes for once or twice weekly in the initial sessions in each phase, after which exercises were done at home *Setting:* clinic for 1/2 sessions then home for the rest of the phase.	Acetaminophen, non-steroidal anti-inflammatory agents and intra-articular injections of glucocorticoids as required. *Applied equally in all treatment groups:* yes
Kirkley 2008	Arthroscopic surgery: saline irrigation of medial, lateral, and	Optimized physical and medical therapy for 12 weeks followed	**A. Physical therapy.** *Description:* not provided. *Frequency intensity and duration:* 1 hour once a week for 12 consecutive weeks. *Supervised:* yes *Setting:* clinic.	Step-wise use of acetaminophen and non-steroidal an-

Table 2. Characteristics of interventions used in included trials *(Continued)*

	patellofemoral joint compartments, based on joint findings one of the following was done - synovectomy; debridement; or excision of degenerative tears of the menisci, fragments of articular cartilage, or chondral flaps and osteophytes that prevented full extension.	by home exercises and arthritis education similar to the exercise group were given post-operatively.	**B. Home exercise along with physical therapy.** *Description:* range-of-motion and strengthening exercises. *Frequency intensity and duration:* twice daily and once on the day of a scheduled physical-therapy session for 12 weeks along with the physical therapy. *Supervised:* no *Setting: h*ome. **C. Unsupervised home exercise.** *Description:* not provided. *Frequency intensity and duration:* duration of the study. *Supervised:* no. *Setting:* home	ti-inflammatory drugs, intra-articular injection of hyaluronic acid and oral glucosamine. Arthritis education - attendance at local Arthritis Society workshops, The Arthritis Helpbook and an educational videotape. *Applied equally in all treatment groups:* yes
Kise 2016	Arthroscopic partial meniscectomy: joint inspection and lavage, probing of menisci and resection of unstable meniscal tissue	Participants were advised to use crutches until normal weight-bearing, and were given written and oral instructions for simple home exercises to be performed two to four times daily.	**Supervised exercise.** *Description:* stationary cycle (20 min) 3 x 10 sets of the following: squat, single-leg squat, step-up, knee stability in pull loop and skating; hamstring on fitball (3 x 8); 2-4 x 15-6 sets of: single-leg leg press, single-leg knee extension, single-leg leg curl; limping cross (3 x 3 rounds). *Frequency intensity and duration:* minimum of two and a maximum of three sessions each week (24 to 36 sessions). Each session lasted approximately 60 to 80 minutes for a total of 12 weeks. *Supervised:* yes *Setting:* clinic	None specified
Osteras 2012	Arthroscopic partial meniscectomy	None specified	**Supervised exercise.** *Description:* 15 to 20 min of aerobic work on a stationary ergometer cycle. After 4 exercises each of 3 sets of 30 repetitions halfway through the exercise program, the participants cycled for 10 min and again after the last 4 exercises, the participants did another 10 min on a stationary ergometer cycle. *Frequency intensity and duration:* 3 times per week for 3 months. *Supervised:* yes. *Setting:* clinic	None specified
Van de Graaf 2018	Arthroscopic partial meniscectomy: standard anteromedial and anterolateral portals were introduced for inspection of the knee joint. The affected meniscus was partially removed until a stable and solid meniscus remained.	Post-operatively, participants received instructions for a home exercise program which consisted of one leg standing for 60 seconds and a step-down exercise comprising 3, 9, 10 repetitions, twice a week.	**Physical therapy (PT).** Participants were referred to PT clinics which were instructed about the exercise protocol by a knee-specialised physical therapist or the primary investigator, prior to the first participant's referral. The PT exercise protocol developed by a knee-specialised physical therapist consisted of 16 sessions of 30 minutes each conducted over 8 weeks. The PT protocol comprised cardiovascular, coordination/balance, and closed kinetic chain strength exercises (in which the distal part of the extremity is fixed to an object that is stationary). If PT failed, the participant was allowed to attend additional PT sessions or have APM, depending on their preference	Home exercise program. *Applied equally in all treatment groups:* yes

Table 2. Characteristics of interventions used in included trials *(Continued)*

Yim 2013	Arthroscopic meniscectomy: meniscal resection with limited debridement of the articular surface lesion.	Post-operatively, all participants were provided with a home exercise program, which was conducted unsupervised, using the same protocol as the non-operative group for 8 weeks.	**A. Supervised exercise.** *Description:* scheduled physical exercise to improve muscle strength, endurance, and flexibility. *Frequency intensity and duration:* 60 minutes per session, 3 times weekly, for 3 weeks. *Supervised:* yes *Setting:* clinic **B. Unsupervised exercise.** *Description:* 3 x 10 sets of the following: half squats with < 45 degrees of flexion with weights, squats with full flexion with weights, knee extension in sitting position, knee flexion in sitting position; stretching of knee extensors and flexors 1 min/muscle group, stationary bicycling (gradual increase every 15 min). *Frequency intensity and duration:* daily for 8 weeks *Supervised:* no *Setting:* home	Analgesics, NSAIDs or muscle relaxants for the first 2 weeks. *Applied equally in all treatment groups:* yes
	Other trials			
Chang 1993	Joint inspection followed by either debridement of torn meniscus and removal of meniscal and cruciate ligament fragments, removal of proliferative synovium, excision of loose articular cartilage fragments, based on the joint findings	Participants were routinely instructed in partial weight-bearing precautions for 10 days post-operatively, followed by physical therapy, consisting of strengthening and flexibility exercises and gait training.	Non-arthroscopic (closed-needle joint) lavage: tidal knee lavage was done under local anaesthesia. A total of 1 litre of saline was injected into and aspirated from the knee in aliquots of 40-120 cc, depending on the size of the knee capsule.	Non-narcotic analgesia and physical therapy, consisting of strengthening and flexibility exercises and gait training. *Applied equally in all treatment groups:* yes
Saeed 2015	Arthroscopic debridement performed using two portals in all cases and under spinal anaesthesia	None reported	Intra-articular hyaluronic acid injections under intradermal anaesthesia given weekly for 5 weeks with a 24-gauge needle under strict aseptic conditions in the operation theatre. In case of joint effusion, aspiration was done before the injection to prevent dilution of the injection.	None specified
Vermesan 2013	Arthroscopic debridement	None reported	A single intra-articular glucocorticoid injection using 1 mL of betamethasone in 4 mL of 1% lidocaine was administered.	None specified
Merchan 1993	Arthroscopic surgery: debridement of synovial tissue; removal of degenerative menisci, osteophytes, and loose bodies; limited debridement of cartilage defects	Post-operatively, a compression bandage was used with early exercises, motion, and weight-bearing as tolerated. Physiotherapy consisting of quadriceps exercises and knee flexion exercises was practiced	The non-operative treatment consisted of non-steroidal anti-inflammatory drugs and a decrease in the intensity of the activities of daily living for a pain-free knee. Physiotherapy was practiced as in the operative group (i.e. quadriceps and knee flexion exercises for 4 weeks).	None specified

Table 2. Characteristics of interventions used in included trials *(Continued)*

	for 4 weeks after surgery.

APM: arthroscopic partial meniscectomy;**ICRS**: International Cartilage Repair Society; **NSAIDS:** non-steroidal anti-inflammatory drugs

Table 3. Outcomes included in analyses

Study ID	Pain	Function	Knee-specific and generic health-related quality of life	Participant-reported treatment success	Knee surgery (replacement or osteotomy)	Serious and total adverse events	Progression of knee OA
Chang 1993	AIMS Pain	AIMS Physical Function	-	Patient global assessment measured on VAS	-	-	-
Gauffin 2014	KOOS Pain	KOOS ADL	KOOS QoL (knee-specific), EQ-5D (generic)	Improvement in the KOOS-Pain score of > 10 points from baseline	-	Yes	Yes
Herrlin 2007	KOOS Pain	KOOS ADL	KOOS QoL (knee-specific)	-	-	-	Yes
Katz 2013	KOOS Pain	WOMAC Physical Function	-	Improvement in the WOMAC-Physical Function score of at least 8 points	Yes	Yes	-
Kirkley 2008	WOMAC Pain	WOMAC Physical Function	Standard-gamble utility score (generic)	-	-	-	-
Kise 2016	KOOS Pain	KOOS ADL	KOOS QoL (knee-specific), SF-36 MCS (generic)	-	Yes	Yes	Yes
Merchan 1993	-	-	-	Increase in post-treatment modified Hospital for Special Surgery Knee Rating Score of at least 10 points	-	Yes	-
Moseley 1996	Average intensity of knee pain (NRS)	-	-	Satisfaction with surgery measured as number of participants reporting 'strongly agree' or 'slightly agree' for item 'do you feel the operation was worthwhile?'	-	-	-
Moseley 2002	SF-36 Pain	SF-36 Physical Function	-	-	-	-[a]	-
Osteras 2012	VAS Pain at rest	KOOS	-	-	-	-	-

Table 3. Outcomes included in analyses *(Continued)*

Roos 2018	KOOS Pain	KOOS ADL	KOOS QoL (knee-specific), SF-36 MCS (generic)	Global perceived effect (rating of 'better' or 'much better')	-	Total adverse events only[a]	-
Saeed 2015	Pain on Knee Society Score System (KSSS)	-	-	-	-	Total adverse events only	-
Sihvonen 2013	Knee pain after exercise (NRS)	Lysholm Knee Score	WOMET score (knee-specific), 15D (generic)	Number of participants reporting 'much better' or 'better' for item 'Is your knee better than before the intervention?	Yes	Yes	Yes
Van de Graaf 2018	VAS Pain on weight-bearing	International Knee Documentation Committee (IKDC) Subjective Knee Form	-	-	Yes	Yes	Yes
Vermesan 2013	-	Oxford Knee Score	-	-	-	-	-
Yim 2013	VAS Pain during activity	Lysholm Knee Score	-	Satisfaction with management, measured as number of participants reporting 'very satisfied' or 'satisfied'	-	-	Yes

[a]Unclear whether serious adverse events occurred in both treatment groups or only in the arthroscopy group
ADL: activities of daily living; **AIMS:** Arthritis Impact Measurement Scale; NRS: numerical rating scale; **EQ-5D 3L:** EuroQoL 5-dimension 3-level quality of life questionnaire; **OA:** osteoarthritis; **KOOS:** Knee injury and Osteoarthritis Outcome Score; **QoL/QOL:** quality of life; **SF-36**: 36-item Short Form Health Survey; **VAS:** visual analogue scale; **WOMAC:** Western Ontario and McMaster Universities Osteoarthritis Index

APPENDICES

Appendix 1. Cochrane Central Register of Controlled Trials (CENTRAL) search strategy

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <April 2021>

Search Strategy:

---

1 exp osteoarthritis/

2 osteoarthr$.tw.

3 (degenerative adj2 arthritis).tw.

4 arthrosis.tw.

5 Menisci, Tibial/

6 menisc$.tw.

7 Arthroscopy/

8 Debridement/

9 (arthroscop$ adj5 debridement).tw.

10 (arthroscop$ adj5 meniscectomy).tw.

11 (arthroscop$ and knee$).tw.

12 1 or 2 or 3 or 4 or 5 or 6

13 7 or 8 or 9 or 10 or 11

14 12 and 13

Appendix 2. MEDLINE search strategy

Database: Ovid MEDLINE(R) <1946 to Present with Daily Update>

Search Strategy:

---

1 exp osteoarthritis/

2 osteoarthr$.tw.

3 (degenerative adj2 arthritis).tw.

4 arthrosis.tw.

5 Menisci, Tibial/

6 menisc$.tw.

7 Arthroscopy/

8 Debridement/

9 (arthroscop$ adj5 debridement).tw.

10 (arthroscop$ adj5 meniscectomy).tw.

11 (arthroscop$ and knee$).tw.

12 1 or 2 or 3 or 4 or 5 or 6

13 7 or 8 or 9 or 10 or 11

14 12 and 13

15 randomized controlled trial.pt.

16 controlled clinical trial.pt.

17 randomized.ab.

18 placebo.ab.

19 drug therapy.fs.

20 randomly.ab.

21 trial.ab.

22 groups.ab.

23 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22

24 exp animals/ not humans.sh.

25 23 not 24

26 14 and 25

Appendix 3. EMBASE search strategy

Database: Embase Classic+Embase <1947 to Present>

Search Strategy:

---

1 osteoarthritis/

2 osteoarthr$.tw.

3 (degenerative adj2 arthritis).tw.

4 arthrosis.tw.

5 exp knee meniscus/

6 menisc$.tw.

7 meniscal repair/

8 exp knee arthroscopy/

9 arthroscopic debridement/

10 (arthroscop$ adj5 debridement).tw.

11 (arthroscop$ adj5 meniscectomy).tw.

12 (arthroscop$ and knee$).tw.

13 or/1-6

14 or/7-12

15 13 and 14

16 random$.tw.

17 factorial$.tw.

18 crossover$.tw.

19 cross over.tw.

20 cross-over.tw.

21 placebo$.tw.

22 (doubl$ adj blind$).tw.

23 (singl$ adj blind$).tw.

24 assign$.tw.

25 allocat$.tw.

26 volunteer$.tw.

27 crossover procedure/

28 double blind procedure/

29 randomized controlled trial/

30 single blind procedure/

31 or/16-30

32 15 and 31

33 limit 32 to exclude medline journals

Appendix 4. Trial registries

ClinicalTrials.Gov

('arthroscopic' or 'arthroscopy' or 'debridement') and ('knee osteoarthritis' or 'meniscal degeneration')

World Health Organization: International Clinical Trials Registry Platform Search Portal

('arthroscopic' or 'arthroscopy' or 'debridement') and ('knee osteoarthritis' or 'meniscal degeneration')

CONTRIBUTIONS OF AUTHORS

RB conceived the review. DOC, RJ and RB drafted the protocol. RJ and SC screened titles and abstracts. RJ, SC and DOC made decisions about inclusion of trials. DOC and SC extracted data. DOC, SC and RJ assessed risk of bias of included trials. DOC and RJ applied GRADE to draw conclusions about the certainty of the body of evidence. SC sought missing data from trial authors and calculated/imputed data if required. DOC analysed data and drafted the review. RB and RJ contributed to writing the review. All authors reviewed and approved the final version.

DECLARATIONS OF INTEREST

Denise O'Connor is an Editor, Renea Johnston a Managing Editor, Sheila Cyril an Assistant Managing Editor and Rachelle Buchbinder the Coordinating Editor with Cochrane Musculoskeletal but they were not involved in editorial decisions regarding this review. Denise O'Connor is also Editor with Cochrane Effective Practice and Organisation of Care (EPOC). They are recipients of an Australian National Health and Medical Research Council (NHMRC) Cochrane Collaboration Round 7 Funding Program Grant, which supports the activities of Cochrane Musculoskeletal - Australia and Cochrane Australia, but the funders do not participate in the conduct of this review. Denise O'Connor is supported by an Australian NHMRC Translating Research into Practice (TRIP) Fellowship (APP1168749). Rachelle Buchbinder is supported by an Australian NHMRC Investigator Grant (APP1194483).

RBP: none known

RWP: none known

POV: none known

SOURCES OF SUPPORT

Internal sources

- Monash Department of Clinical Epidemiology, Cabrini Institute, Australia

  In kind support
- Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Australia

  In kind

External sources

- National Health and Medical Research Council (NHMRC), Australia

  Denise O'Connor and Rachelle Buchbinder are supported by NHMRC fellowships

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We extracted end-of-treatment mean (SD) for pain, function, quality of life (QoL), used SMD as scales differed, then back-translated using SD from control at baseline (in comparison, the earlier review converted all to change-from-baseline using a common scale, then calculated difference in change scores between groups, and used MD).

Our planned outcome was subsequent knee replacement but we changed it to subsequent knee surgery (including knee replacement and any other surgery to treat severe knee osteoarthritis; e.g. high tibial osteotomy).

INDEX TERMS

Medical Subject Headings (MeSH)

*Arthroscopy [adverse effects]; *Osteoarthritis, Knee [surgery]; Pain Measurement; Pain, Postoperative; Quality of Life

MeSH check words

Aged; Female; Humans; Middle Aged